

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 18, 2004, 04:20:54 ; Search time 9.88235 Seconds  
(without alignments)  
31.344 Million cell updates/sec

Title: US-10-615-959-39

Perfect score: 32

Sequence: 1 EKAGGF 6

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents\_AA.\*

1: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pep.\*

2: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep.\*

3: /cgn2\_6/ptodata/2/iaa/6A\_COMB.pep.\*

4: /cgn2\_6/ptodata/2/iaa/6B\_COMB.pep.\*

5: /cgn2\_6/ptodata/2/iaa/PCTUS\_COMB.pep.\*

6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	32	100.0	6	4	US-10-009-999A-39
2	32	100.0	7	3	US-08-963-825-15
3	32	100.0	7	4	US-09-500-811-15
4	32	100.0	7	4	US-09-570-573-15
5	32	100.0	7	4	US-09-548-608-15
6	32	100.0	9	4	US-10-009-999A-41
7	32	100.0	12	4	US-10-009-999A-43
8	32	100.0	15	4	US-08-278-774-20
9	32	100.0	623	3	US-09-029-348-3
10	32	100.0	626	3	US-09-029-348-2
11	32	100.0	1078	3	US-08-963-825-21
12	32	100.0	1078	4	US-09-500-811-21
13	32	100.0	1078	4	US-09-570-573-21
14	32	100.0	1078	4	US-09-548-608-21
15	28	87.5	301	4	US-09-252-991A-33096
16	28	87.5	725	4	US-09-252-991A-27825
17	28	87.5	750	4	US-09-489-039A-8940
18	28	87.5	752	4	US-09-252-991A-21724
19	27	84.4	53	3	US-08-905-223-489
20	27	84.4	70	4	US-09-621-976-6612
21	27	84.4	72	4	US-09-621-976-5811
22	27	84.4	131	4	US-09-621-976-4204
23	27	84.4	168	4	US-09-543-681A-6129
24	27	84.4	182	4	US-09-489-039A-13197
25	27	84.4	219	4	US-09-107-532A-3911
26	27	84.4	250	4	US-09-540-236-3600
27	27	84.4	275	4	US-09-252-991A-30551

Sequence 3826, Ap  
Sequence 4213, Ap  
Sequence 3392, Ap  
Sequence 32, Appl  
Sequence 9821, Ap  
Sequence 6580, Ap  
Sequence 6678, Ap  
Sequence 18710, A  
Sequence 8, Appl  
Sequence 5, Appl  
Sequence 5, Appl  
Sequence 2, Appl  
Sequence 2, Appl  
Sequence 8778, Ap  
Sequence 28000, A  
Sequence 13405, A  
Sequence 10684, A

## ALIGNMENTS

### RESULT 1

US-10-009-999A-39  
; Sequence 39, Application US/10009999A  
; Patent No. 6602980  
; GENERAL INFORMATION:  
; APPLICANT: Eyre, David R.  
; TITLE OF INVENTION: SYNTHETIC PEPTIDES FOR COLLAGEN RESORPTION ASSAYS  
; FILE REFERENCE: WROS-1-18220  
; CURRENT APPLICATION NUMBER: US/10/009,999A  
; CURRENT FILING DATE: 2003-01-22  
; PRIOR APPLICATION NUMBER: PCT/US99/29357  
; PRIOR FILING DATE: 1999-12-10  
; PRIOR APPLICATION NUMBER: US 09/335,098  
; PRIOR FILING DATE: 1999-06-17  
; PRIOR APPLICATION NUMBER: US 60/141,574  
; PRIOR FILING DATE: 1999-06-29  
; PRIOR APPLICATION NUMBER: US 60/142,274  
; PRIOR FILING DATE: 1999-07-02  
; PRIOR APPLICATION NUMBER: US 60/142,675  
; PRIOR FILING DATE: 1999-07-07  
; PRIOR APPLICATION NUMBER: US 09/385,740  
; PRIOR FILING DATE: 1999-08-30  
; NUMBER OF SEQ ID NOS: 45  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 39  
; LENGTH: 6  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: MISC FEATURE  
; OTHER INFORMATION: synthetic peptide corresponding to carboxy-terminal  
; OTHER INFORMATION: telopeptide sequence of human type III collagen  
US-10-009-999A-39

Query Match 100.0%; Score 32; DB 4; Length 6;

Best Local Similarity 100.0%; Pred. No. 3e+05;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EKAGGF 6

DB 1 EKAGGF 6

### RESULT 2

US-08-963-825-15

; Sequence 15, Application US/08963825

; Patent No. 6110689

; GENERAL INFORMATION:

; APPLICANT: Oviatt, Per

; APPLICANT: Bonde, Martin

```
;
; TITLE OF INVENTION: A Method for Assaying Collagen Fragments
; TITLE OF INVENTION: in Body Fluids, A Test Kit and Means for Carrying Out the
; TITLE OF INVENTION: Method and Use of the Method to Diagnose the Presence of
; TITLE OF INVENTION: Disorders Associated with the Metabolism of
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Darby & Darby PC
; STREET: 805 Third Avenue
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10022
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/963,825
; FILING DATE:
; CLASSIFICATION: 436
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/187,319
; FILING DATE: 21-JAN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Gogoris, Adda C
; REGISTRATION NUMBER: 29,714
; REFERENCE/DOCKET NUMBER: 4305/08701
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-527-7700
; TELEFAX: 212-753-6237
; TELEX: 236687
;
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FRAGMENT TYPE: C-terminal
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; IMMEDIATE SOURCE:
; CLONE: COLLAGEN TYPE III - ALPHA 1 - C TERM
;
; US-08-963-825-15
;
; Query Match 100.0%; Score 32; DB 3; Length 7;
; Best Local Similarity 100.0%; Pred. No. 3e+05;
; Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
; QY 1 EKAGGF 6
; | | | | |
; Db 1 EKAGGF 6
;
; RESULT 3
; US-09-500-811-15
; Sequence 15, Application US/09500811
; Patent No. 6323314
; GENERAL INFORMATION:
; APPLICANT: Qvist, Per
; APPLICANT: Bonde, Martin
; TITLE OF INVENTION: A Method for Assaying Collagen Fragments
; TITLE OF INVENTION: in Body Fluids, A Test Kit and Means for Carrying Out the
; TITLE OF INVENTION: Method and Use of the Method to Diagnose the Presence of
; TITLE OF INVENTION: Disorders Associated with the Metabolism of
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Darby & Darby PC
; STREET: 805 Third Avenue
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10022
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/570,573
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/187,319
; FILING DATE:
;
; US-09-570-573-15
; Sequence 15, Application US/09570573
; Patent No. 6342361
; GENERAL INFORMATION:
; APPLICANT: Qvist, Per
; APPLICANT: Bonde, Martin
; TITLE OF INVENTION: A Method for Assaying Collagen Fragments
; TITLE OF INVENTION: in Body Fluids, A Test Kit and Means for Carrying Out the
; TITLE OF INVENTION: Method and Use of the Method to Diagnose the Presence of
; TITLE OF INVENTION: Disorders Associated with the Metabolism of
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Darby & Darby PC
; STREET: 805 Third Avenue
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10022
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/570,573
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/187,319
; FILING DATE:
```

```
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/500,811
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/187,319
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Gogoris, Adda C
; REGISTRATION NUMBER: 29,714
; REFERENCE/DOCKET NUMBER: 4305/08701
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-527-7700
; TELEFAX: 212-753-6237
; TELEX: 236687
;
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FRAGMENT TYPE: C-terminal
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; IMMEDIATE SOURCE:
; CLONE: COLLAGEN TYPE III - ALPHA 1 - C TERM
;
; US-09-500-811-15
;
; Query Match 100.0%; Score 32; DB 4; Length 7;
; Best Local Similarity 100.0%; Pred. No. 3e+05;
; Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
; QY 1 EKAGGF 6
; | | | | |
; Db 1 EKAGGF 6
;
; RESULT 4
; US-09-570-573-15
; Sequence 15, Application US/09570573
; Patent No. 6342361
; GENERAL INFORMATION:
; APPLICANT: Qvist, Per
; APPLICANT: Bonde, Martin
; TITLE OF INVENTION: A Method for Assaying Collagen Fragments
; TITLE OF INVENTION: in Body Fluids, A Test Kit and Means for Carrying Out the
; TITLE OF INVENTION: Method and Use of the Method to Diagnose the Presence of
; TITLE OF INVENTION: Disorders Associated with the Metabolism of
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Darby & Darby PC
; STREET: 805 Third Avenue
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10022
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/570,573
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/187,319
; FILING DATE:
```



```
; ATTORNEY/AGENT INFORMATION:
; NAME: Gogoris, Adda C
; REGISTRATION NUMBER: 29,714
; REFERENCE/DOCKET NUMBER: 4305/08701
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-527-7700
; TELEFAX: 212-753-6237
; TELEX: 236687
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FRAGMENT TYPE: C-terminal
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; IMMEDIATE SOURCE:
; CLONE: COLLAGEN TYPE III - ALPHA 1 - C TERM
US-09-548-608-15
Query Match 100.0%; Score 32; DB 4; Length 7;
Best Local Similarity 100.0%; Pred. No. 3e+05;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 EKAGGF 6
DB 1 EKAGGF 6
RESULT 5
US-09-548-608-15
; Sequence 15, Application US/09548608
; Patent No. 6355442
; GENERAL INFORMATION:
; APPLICANT: Ovist, Per
; APPLICANT: Bonde, Martin
; TITLE OF INVENTION: A Method for Assaying Collagen Fragments
; TITLE OF INVENTION: in Body Fluids, A Test Kit and Means for Carrying Out the
; TITLE OF INVENTION: Method and Use of the Method to Diagnose the Presence of
; TITLE OF INVENTION: Disorders Associated with the Metabolism of
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Darby & Darby PC
; STREET: 805 Third Avenue
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/548,608
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/187,319
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Gogoris, Adda C
; REGISTRATION NUMBER: 29,714
; REFERENCE/DOCKET NUMBER: 4305/08701
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-527-7700
; TELEFAX: 212-753-6237
; TELEX: 236687
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FRAGMENT TYPE: C-terminal
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; IMMEDIATE SOURCE:
; CLONE: COLLAGEN TYPE III - ALPHA 1 - C TERM
US-09-548-608-15
Query Match 100.0%; Score 32; DB 4; Length 7;
Best Local Similarity 100.0%; Pred. No. 3e+05;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 EKAGGF 6
DB 1 EKAGGF 6
RESULT 6
US-10-009-999A-41
; Sequence 41, Application US/10009999A
; Patent No. 6602980
; GENERAL INFORMATION:
; APPLICANT: Eyre, David R.
; TITLE OF INVENTION: SYNTHETIC PEPTIDES FOR COLLAGEN RESORPTION ASSAYS
; FILE REFERENCE: WROS-1-18220
; CURRENT APPLICATION NUMBER: US/10/009,999A
; CURRENT FILING DATE: 2003-01-22
; PRIOR APPLICATION NUMBER: PCT/US99/29357
; PRIOR FILING DATE: 1999-12-10
; PRIOR APPLICATION NUMBER: US 09/335,098
; PRIOR FILING DATE: 1999-06-17
; PRIOR APPLICATION NUMBER: US 60/141,574
; PRIOR FILING DATE: 1999-06-29
; PRIOR APPLICATION NUMBER: US 60/142,274
; PRIOR FILING DATE: 1999-07-02
; PRIOR APPLICATION NUMBER: US 60/142,675
; PRIOR FILING DATE: 1999-07-07
; PRIOR APPLICATION NUMBER: US 09/385,740
; PRIOR FILING DATE: 1999-08-30
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 41
; LENGTH: 9
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: MISC FEATURE
; OTHER INFORMATION: synthetic peptide corresponding to carboxy-terminal
; OTHER INFORMATION: telopeptide sequence of human type III collagen
US-10-009-999A-41
Query Match 100.0%; Score 32; DB 4; Length 9;
Best Local Similarity 100.0%; Pred. No. 3e+05;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 EKAGGF 6
DB 4 EKAGGF 9
RESULT 7
US-10-009-999A-43
; Sequence 43, Application US/10009999A
; Patent No. 6602980
; GENERAL INFORMATION:
; APPLICANT: Eyre, David R.
; TITLE OF INVENTION: SYNTHETIC PEPTIDES FOR COLLAGEN RESORPTION ASSAYS
; FILE REFERENCE: WROS-1-18220
; CURRENT APPLICATION NUMBER: US/10/009,999A
; CURRENT FILING DATE: 2003-01-22
; PRIOR APPLICATION NUMBER: PCT/US99/29357
; PRIOR FILING DATE: 1999-12-10
```

```
; PRIOR APPLICATION NUMBER: US 09/335,098
; PRIOR FILING DATE: 1999-06-17
; PRIOR APPLICATION NUMBER: US 60/141,574
; PRIOR FILING DATE: 1999-06-29
; PRIOR APPLICATION NUMBER: US 60/142,274
; PRIOR FILING DATE: 1999-07-02
; PRIOR APPLICATION NUMBER: US 60/142,675
; PRIOR FILING DATE: 1999-07-07
; PRIOR APPLICATION NUMBER: US 09/385,740
; PRIOR FILING DATE: 1999-08-30
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 43
; LENGTH: 12
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: MISC FEATURE
; OTHER INFORMATION: synthetic peptide corresponding to carboxy-terminal
; OTHER INFORMATION: telopeptide sequence of human type III collagen
US-10-009-999A-43

Query Match      100.0%; Score 32; DB 4; Length 12;
Best Local Similarity 100.0%; Pred. No. 2;
Matches      6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 EKAGGF 6
Db      7 EKAGGF 12

RESULT 8
US-08-278-774-20
; Sequence 20, Application US/08278774
; Patent No. 6653450
; GENERAL INFORMATION:
; APPLICANT: Berg, Richard A
; APPLICANT: Toman, David P
; APPLICANT: Wallace, Donald
; TITLE OF INVENTION: MUTATED RECOMBINANT COLLAGENS
; NUMBER OF SEQUENCES: 33
; CORRESPONDENCE ADDRESS:
; ADDRESS: COLLAGEN CORPORATION
; STREET: 2500 Faber Place
; CITY: Palo Alto
; STATE: California
; COUNTRY: USA
; ZIP: 94303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/278,774
; FILING DATE: 22-JUL-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Rafayko, Kathi L
; REGISTRATION NUMBER: 36,644
; REFERENCE/DOCKET NUMBER: 94-018
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 354-4642
; TELEFAX: (415) 354-4752
; TELEX:
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 15 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-278-774-20
```

```
Query Match      100.0%; Score 32; DB 4; Length 15;
Best Local Similarity 100.0%; Pred. No. 2.5;
Matches      6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 EKAGGF 6
Db      5 EKAGGF 10

RESULT 9
US-09-029-348-3
; Sequence 3, Application US/09029348
; Patent No. 6171827
; GENERAL INFORMATION:
; APPLICANT: THE VICTORIA UNIVERSITY OF MANCHESTER
; TITLE OF INVENTION: NOVEL PROCOLLAGENS
; FILE REFERENCE: G087857PUS LISTING
; CURRENT APPLICATION NUMBER: US/09/029,348
; CURRENT FILING DATE: 1998-05-07
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn ver. 2.0
; SEQ ID NO 3
; LENGTH: 623
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: SEQUENCE
; OTHER INFORMATION: DERIVED FROM CDNA OF PROCOLLAGENS
US-09-029-348-3
```

```
Query Match      100.0%; Score 32; DB 3; Length 623;
Best Local Similarity 100.0%; Pred. No. 94;
Matches      6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 EKAGGF 6
Db      371 EKAGGF 376
```

```
RESULT 10
US-09-029-348-2
; Sequence 2, Application US/09029348
; Patent No. 6171827
; GENERAL INFORMATION:
; APPLICANT: THE VICTORIA UNIVERSITY OF MANCHESTER
; TITLE OF INVENTION: NOVEL PROCOLLAGENS
; FILE REFERENCE: G087857PUS LISTING
; CURRENT APPLICATION NUMBER: US/09/029,348
; CURRENT FILING DATE: 1998-05-07
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn ver. 2.0
; SEQ ID NO 2
; LENGTH: 626
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: SEQUENCE
; OTHER INFORMATION: DERIVED FROM CDNA OF PROCOLLAGENS
US-09-029-348-2
```

```
Query Match      100.0%; Score 32; DB 3; Length 626;
Best Local Similarity 100.0%; Pred. No. 95;
Matches      6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 EKAGGF 6
Db      371 EKAGGF 376
```

```
RESULT 11
US-08-963-825-21
; Sequence 21, Application US/08963825
```

```

; Patent No. 6110689
; GENERAL INFORMATION:
; APPLICANT: Qvist, Per
; TITLE OF INVENTION: A Method for Assaying Collagen Fragments
; TITLE OF INVENTION: in Body Fluids, A Test Kit and Means for Carrying Out the
; TITLE OF INVENTION: Method and Use of the Method to Diagnose the Presence of
; TITLE OF INVENTION: Disorders Associated with the Metabolism of
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Darby & Darby PC
; STREET: 805 Third Avenue
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/963,825
; FILING DATE:
; CLASSIFICATION: 436
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/187,319
; FILING DATE: 21-JAN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Gogoris, Adda C
; REGISTRATION NUMBER: 29,714
; REFERENCE/DOCKET NUMBER: 4305/08701
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-527-7700
; TELEFAX: 212-753-6237
; TELEX: 236687
; INFORMATION FOR SEQ ID NO: 21:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1078 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; IMMEDIATE SOURCE:
; CLONE: COLLAGEN ALPHA 1 (III)
; US-08-963-825-21

Query Match 100.0%; Score 32; DB 3; Length 1078;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EKAGGF 6
Db 1064 EKAGGF 1069

RESULT 12
US-09-500-811-21
; Sequence 21, Application US/09500811
; Patent No. 6323314
; GENERAL INFORMATION:
; APPLICANT: Qvist, Per
; TITLE OF INVENTION: A Method for Assaying Collagen Fragments
; TITLE OF INVENTION: in Body Fluids, A Test Kit and Means for Carrying Out the
; TITLE OF INVENTION: Method and Use of the Method to Diagnose the Presence of
; TITLE OF INVENTION: Disorders Associated with the Metabolism of
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Darby & Darby PC
; STREET: 805 Third Avenue
; CITY: New York

```

```

; STATE: New York
; COUNTRY: USA
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/500,811
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/187,319
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Gogoris, Adda C
; REGISTRATION NUMBER: 29,714
; REFERENCE/DOCKET NUMBER: 4305/08701
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-527-7700
; TELEFAX: 212-753-6237
; TELEX: 236687
; INFORMATION FOR SEQ ID NO: 21:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1078 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; IMMEDIATE SOURCE:
; CLONE: COLLAGEN ALPHA 1 (III)
; US-09-500-811-21

Query Match 100.0%; Score 32; DB 4; Length 1078;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EKAGGF 6
Db 1064 EKAGGF 1069

RESULT 13
US-09-570-573-21
; Sequence 21, Application US/09570573
; Patent No. 6342361
; GENERAL INFORMATION:
; APPLICANT: Qvist, Per
; TITLE OF INVENTION: A Method for Assaying Collagen Fragments
; TITLE OF INVENTION: in Body Fluids, A Test Kit and Means for Carrying Out the
; TITLE OF INVENTION: Method and Use of the Method to Diagnose the Presence of
; TITLE OF INVENTION: Disorders Associated with the Metabolism of
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Darby & Darby PC
; STREET: 805 Third Avenue
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/570,573
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:

```

```
; APPLICATION NUMBER: 08/187,319
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Gogoris, Adda C
; REGISTRATION NUMBER: 29,714
; REFERENCE/DOCKET NUMBER: 4305/08701
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-527-7700
; TELEFAX: 212-753-6237
; TELEX: 236687
; INFORMATION FOR SEQ ID NO: 21:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1078 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; IMMEDIATE SOURCE:
; CLONE: COLLAGEN ALPHA 1 (III)
US-09-570-573-21

Query Match 100.0%; Score 32; DB 4; Length 1078;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 EKAGGF 6
Db 1064 EKAGGF 1069

RESULT 14
US-09-548-608-21
; Sequence 21, Application US/09548608
; Patent No. 6355442
; GENERAL INFORMATION:
; APPLICANT: Qvist, Per
; APPLICANT: Bonde, Martin
; TITLE OF INVENTION: A Method for Assaying Collagen Fragments
; TITLE OF INVENTION: In Body Fluids, A Test Kit and Means for Carrying Out the
; TITLE OF INVENTION: Method and Use of the Method to Diagnose the Presence of
; TITLE OF INVENTION: Disorders Associated with the Metabolism of
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Darby & Darby PC
; STREET: 805 Third Avenue
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/548,608
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/187,319
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Gogoris, Adda C
; REGISTRATION NUMBER: 29,714
; REFERENCE/DOCKET NUMBER: 4305/08701
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-527-7700
; TELEFAX: 212-753-6237
; TELEX: 236687
; INFORMATION FOR SEQ ID NO: 21:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1078 amino acids
```

```
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; IMMEDIATE SOURCE:
; CLONE: COLLAGEN ALPHA 1 (III)
US-09-548-608-21

Query Match 100.0%; Score 32; DB 4; Length 1078;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 EKAGGF 6
Db 1064 EKAGGF 1069

RESULT 15
US-09-252-991A-33096
; Sequence 33096, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 33096
; LENGTH: 301
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-33096

Query Match 87.5%; Score 28; DB 4; Length 301;
Best Local Similarity 83.3%; Pred. No. 2.9e+02;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 EKAGGF 6
Db 184 EKAGGF 189

Search completed: September 18, 2004, 04:32:20
Job time : 10.8824 secs
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 18, 2004, 04:29:56 ; Search time 28.4706 Seconds  
(without alignments)  
67.677 Million cell updates/sec

Title: US-10-615-959-39

Perfect score: 32

Sequence: 1 EKAGGF 6

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1342398 seqs, 321133274 residues

Total number of hits satisfying chosen parameters: 1342398

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

#### Database :

Published Applications AA:\*

- 1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*
- 2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep.\*
- 3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*
- 4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*
- 5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*
- 6: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep.\*
- 7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*
- 8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*
- 9: /cgn2\_6/ptodata/2/pubpaa/US09A\_PUBCOMB.pep.\*
- 10: /cgn2\_6/ptodata/2/pubpaa/US09B\_PUBCOMB.pep.\*
- 11: /cgn2\_6/ptodata/2/pubpaa/US09C\_PUBCOMB.pep.\*
- 12: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep.\*
- 13: /cgn2\_6/ptodata/2/pubpaa/US10A\_PUBCOMB.pep.\*
- 14: /cgn2\_6/ptodata/2/pubpaa/US10B\_PUBCOMB.pep.\*
- 15: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pep.\*
- 16: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep.\*
- 17: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*
- 18: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	32	100.0	6	12	US-10-615-959-39
2	32	100.0	7	14	US-10-058-124-15
3	32	100.0	9	12	US-10-615-959-41
4	32	100.0	12	12	US-10-615-959-43
5	32	100.0	25	15	US-10-366-125-11
6	32	100.0	51	11	US-09-864-408A-5882
7	32	100.0	166	16	US-10-767-701-33455
8	32	100.0	262	15	US-10-366-125-19
9	32	100.0	309	9	US-09-908-711-78
10	32	100.0	486	16	US-10-437-963-162979
11	32	100.0	1078	14	US-10-058-124-21
12	32	100.0	1466	12	US-09-918-715-226
13	32	100.0	1466	12	US-10-257-021-72
14	32	100.0	1466	14	US-10-177-293-68
15	32	100.0	1466	14	US-10-301-822-33

16	32	100.0	1466	15	US-10-402-089-4	Sequence 4, Appli
17	32	100.0	1466	15	US-10-402-089-6	Sequence 6, Appli
18	32	100.0	1466	15	US-10-402-089-12	Sequence 12, Appli
19	32	100.0	1466	15	US-10-402-072A-4	Sequence 4, Appli
20	32	100.0	1466	15	US-10-402-072A-6	Sequence 6, Appli
21	32	100.0	1466	15	US-10-402-072A-12	Sequence 12, Appli
22	32	100.0	1466	16	US-10-357-851-3	Sequence 3, Appli
23	32	100.0	1466	16	US-10-358-024-3	Sequence 3, Appli
24	32	100.0	1466	16	US-10-734-564-103	Sequence 103, App
25	32	100.0	1687	13	US-10-094-679-3	Sequence 3, Appli
26	29	90.6	68	16	US-10-437-963-151038	Sequence 151038,
27	29	90.6	85	12	US-10-425-114-46873	Sequence 46873, A
28	29	90.6	90	16	US-10-437-963-131724	Sequence 131724,
29	29	90.6	99	16	US-10-437-963-202123	Sequence 202123,
30	29	90.6	101	16	US-10-437-963-132375	Sequence 132375,
31	29	90.6	213	12	US-10-425-114-52424	Sequence 52424, A
32	29	90.6	229	12	US-10-424-599-261871	Sequence 261871,
33	29	90.6	362	12	US-10-425-114-39370	Sequence 39370, A
34	29	90.6	372	14	US-10-156-761-10310	Sequence 10310, A
35	29	90.6	387	12	US-10-424-599-227821	Sequence 227821,
36	29	90.6	389	16	US-10-408-765A-2121	Sequence 2121, Ap
37	29	90.6	392	15	US-10-310-154-730	Sequence 730, App
38	29	90.6	398	9	US-09-925-301-1429	Sequence 1429, Ap
39	29	90.6	410	16	US-10-408-765A-749	Sequence 749, App
40	29	90.6	417	10	US-09-935-642-10	Sequence 10, Appl
41	29	90.6	417	12	US-10-221-278-228	Sequence 228, App
42	29	90.6	417	12	US-10-170-385-255	Sequence 255, App
43	29	90.6	417	14	US-10-316-253-302	Sequence 302, App
44	29	90.6	417	14	US-10-316-253-304	Sequence 304, App
45	29	90.6	417	15	US-10-291-172-228	Sequence 228, App

#### ALIGNMENTS

#### RESULT 1

US-10-615-959-39  
; Sequence 39, Application US/10615959  
; Publication No. US20040048321A1  
; GENERAL INFORMATION:  
; APPLICANT: Eyre, David R.  
; TITLE OF INVENTION: SYNTHETIC PEPTIDES FOR COLLAGEN RESORPTION ASSAYS  
; FILE REFERENCE: WROS-1-18220  
; CURRENT APPLICATION NUMBER: US/10/615,959  
; CURRENT FILING DATE: 2003-07-08  
; PRIOR APPLICATION NUMBER: US/10/009,999A  
; PRIOR FILING DATE: 2003-01-22  
; PRIOR APPLICATION NUMBER: PCT/US99/29357  
; PRIOR FILING DATE: 1999-12-10  
; PRIOR APPLICATION NUMBER: US 09/335,098  
; PRIOR FILING DATE: 1999-06-17  
; PRIOR APPLICATION NUMBER: US 60/141,574  
; PRIOR FILING DATE: 1999-06-29  
; PRIOR APPLICATION NUMBER: US 60/142,274  
; PRIOR FILING DATE: 1999-07-02  
; PRIOR APPLICATION NUMBER: US 60/142,675  
; PRIOR FILING DATE: 1999-07-07  
; PRIOR APPLICATION NUMBER: US 09/385,740  
; NUMBER OF SEQ ID NOS: 45  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 39  
; LENGTH: 6  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; NAME/KEY: MISC\_FEATURE  
; OTHER INFORMATION: synthetic peptide corresponding to carboxy-terminal  
; OTHER INFORMATION: telopeptide sequence of human type III collagen  
US-10-615-959-39

Query Match 100.0%; Score 32; DB 12; Length 6;  
Best Local Similarity 100.0%; Pred. No. 1.2e+06;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 EKAGGF 6  
Db 1 EKAGGF 6

RESULT 2  
US-10-058-124-15  
Sequence 15, Application US/10058124  
Publication No. US20030119058A1  
GENERAL INFORMATION:  
APPLICANT: Qvist, Per  
Bonde, Martin  
TITLE OF INVENTION: A Method for Assaying Collagen Fragments  
in Body Fluids, A Test Kit and Means for Carrying Out the  
Method and Use of the Method to Diagnose the Presence of  
Disorders Associated with the Metabolism of  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Darby & Darby PC  
STREET: 805 Third Avenue  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10022  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/058.124  
FILING DATE: 29-Jan-2002  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 09/570,573  
FILING DATE: 2002-MAY-12  
APPLICATION NUMBER: 08/187,319  
FILING DATE: <Unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: Gogoris, Adda C.  
REGISTRATION NUMBER: 29,714  
REFERENCE/DOCKET NUMBER: 4305/08701  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-527-7700  
TELEFAX: 212-753-6237  
TELEX: 236687  
INFORMATION FOR SEQ ID NO: 15:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 7 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: Peptide  
FRAGMENT TYPE: C-terminal  
ORIGINAL SOURCE:  
ORGANISM: Homo sapiens  
CLONE: COLLAGEN TYPE III - ALPHA 1 - C TERM  
SEQUENCE DESCRIPTION: SEQ ID NO: 15:  
US-10-058-124-15  
Query Match 100.0%; Score 32; DB 14; Length 7;  
Best Local Similarity 100.0%; Pred. No. 1.2e+06;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 EKAGGF 6  
Db 1 EKAGGF 6

RESULT 3  
US-10-615-959-41  
Sequence 41, Application US/10615959  
Publication No. US20040048321A1  
GENERAL INFORMATION:  
APPLICANT: Eyre, David R.  
TITLE OF INVENTION: SYNTHETIC PEPTIDES FOR COLLAGEN RESORPTION ASSAYS  
FILE REFERENCE: WROS-1-18220  
CURRENT APPLICATION NUMBER: US/10/615,959  
PRIOR FILING DATE: 2003-07-08  
PRIOR APPLICATION NUMBER: US/10/009,999A  
PRIOR FILING DATE: 2003-01-22  
PRIOR APPLICATION NUMBER: PCT/US99/29357  
PRIOR FILING DATE: 1999-12-10  
PRIOR APPLICATION NUMBER: US 09/335,098  
PRIOR FILING DATE: 1999-06-17  
PRIOR APPLICATION NUMBER: US 60/141,574  
PRIOR FILING DATE: 1999-06-29  
PRIOR APPLICATION NUMBER: US 60/142,274  
PRIOR FILING DATE: 1999-07-02  
PRIOR APPLICATION NUMBER: US 60/142,675  
PRIOR FILING DATE: 1999-07-07  
PRIOR APPLICATION NUMBER: US 09/385,740  
PRIOR FILING DATE: 1999-08-30  
NUMBER OF SEQ ID NOS: 45  
SOFTWARE: Patent In version 3.2  
SEQ ID NO 41  
LENGTH: 9  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: MISC FEATURE  
OTHER INFORMATION: synthetic peptide corresponding to carboxy-terminal  
telopeptide sequence of human type III collagen  
US-10-615-959-41  
Query Match 100.0%; Score 32; DB 12; Length 9;  
Best Local Similarity 100.0%; Pred. No. 1.2e+06;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 EKAGGF 6  
Db 4 EKAGGF 9

RESULT 4  
US-10-615-959-43  
Sequence 43, Application US/10615959  
Publication No. US20040048321A1  
GENERAL INFORMATION:  
APPLICANT: Eyre, David R.  
TITLE OF INVENTION: SYNTHETIC PEPTIDES FOR COLLAGEN RESORPTION ASSAYS  
FILE REFERENCE: WROS-1-18220  
CURRENT APPLICATION NUMBER: US/10/615,959  
CURRENT FILING DATE: 2003-07-08  
PRIOR APPLICATION NUMBER: US/10/009,999A  
PRIOR FILING DATE: 2003-01-22  
PRIOR APPLICATION NUMBER: PCT/US99/29357  
PRIOR FILING DATE: 1999-12-10  
PRIOR APPLICATION NUMBER: US 09/335,098  
PRIOR FILING DATE: 1999-06-17  
PRIOR APPLICATION NUMBER: US 60/141,574  
PRIOR FILING DATE: 1999-06-29  
PRIOR APPLICATION NUMBER: US 60/142,274  
PRIOR FILING DATE: 1999-07-02  
PRIOR APPLICATION NUMBER: US 60/142,675  
PRIOR FILING DATE: 1999-07-07  
PRIOR APPLICATION NUMBER: US 09/385,740  
PRIOR FILING DATE: 1999-08-30  
NUMBER OF SEQ ID NOS: 45  
SOFTWARE: Patent In version 3.2  
SEQ ID NO 43  
LENGTH: 12  
TYPE: PRT  
ORGANISM: Homo sapiens

```
;
; NAME/KEY: MISC FEATURE
; OTHER INFORMATION: synthetic peptide corresponding to carboxy-terminal
; OTHER INFORMATION: telopeptide sequence of human type III collagen
US-10-615-959-43

Query Match      100.0%; Score 32; DB 12; Length 12;
Best Local Similarity 100.0%; Pred. No. 7.5; Indels 0; Gaps 0;
Matches 6; Conservative 0; Mismatches 0;

QY 1 EKAGGF 6
Db 7 EKAGGF 12

RESULT 5
US-10-366-125-11
; Sequence 11, Application US/10366125
; Publication No. US20030228259A1
; GENERAL INFORMATION:
; APPLICANT: Hellerstein, Marc
; TITLE OF INVENTION: MEASUREMENT OF BIOSYNTHESIS AND BREAKDOWN RATES OF
; TITLE OF INVENTION: BIOLOGICAL MOLECULES THAT ARE INACCESSIBLE OR NOT
; TITLE OF INVENTION: EASILY ACCESSIBLE TO DIRECT SAMPLING, NON-INVASIVELY,
; TITLE OF INVENTION: BY LABEL INCORPORATION INTO METABOLIC DERIVATIVES AND
; TITLE OF INVENTION: CATABOLIC PRODUCTS
; FILE REFERENCE: 416272003500
; CURRENT APPLICATION NUMBER: US/10/366,125
; CURRENT FILING DATE: 2003-02-12
; PRIOR APPLICATION NUMBER: US 60/356,008
; PRIOR FILING DATE: 2002-02-12
; NUMBER OF SEQ ID NOS: 28
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 25
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-366-125-11

Query Match      100.0%; Score 32; DB 15; Length 25;
Best Local Similarity 100.0%; Pred. No. 16; Indels 0; Gaps 0;
Matches 6; Conservative 0; Mismatches 0;

QY 1 EKAGGF 6
Db 15 EKAGGF 20

RESULT 6
US-09-864-408A-5882
; Sequence 5882, Application US/09864408A
; Publication No. US20040009474A1
; GENERAL INFORMATION:
; APPLICANT: Leach, Martin D.
; APPLICANT: Shimkets, Richard A.
; TITLE OF INVENTION: No. US20040009474A1 Human Polynucleotides and Polypeptides Encod
; FILE REFERENCE: 21402-012
; CURRENT APPLICATION NUMBER: US/09/864,408A
; CURRENT FILING DATE: 2001-05-24
; PRIOR APPLICATION NUMBER: 60/206,690
; PRIOR FILING DATE: 2000-05-24
; NUMBER OF SEQ ID NOS: 9068
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 582
; LENGTH: 51
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-864-408A-5882

Query Match      100.0%; Score 32; DB 11; Length 51;
Best Local Similarity 100.0%; Pred. No. 32; Indels 0; Gaps 0;
Matches 6; Conservative 0; Mismatches 0;
```

```
QY 1 EKAGGF 6
Db 28 EKAGGF 33

RESULT 7
US-10-767-701-33455
; Sequence 33455, Application US/10767701
; Publication No. US20040172684A1
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants and Uses Thereof For Plant Improvement
; FILE REFERENCE: 38-21(53535)B
; CURRENT APPLICATION NUMBER: US/10/767,701
; CURRENT FILING DATE: 2004-01-29
; NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 33455
; LENGTH: 166
; TYPE: PRT
; ORGANISM: Sorghum bicolor
; FEATURE:
; OTHER INFORMATION: Clone ID: SORBI-28MAY03-C32355_1.pep
US-10-767-701-33455

Query Match      100.0%; Score 32; DB 16; Length 166;
Best Local Similarity 100.0%; Pred. No. 1e+02; Indels 0; Gaps 0;
Matches 6; Conservative 0; Mismatches 0;

QY 1 EKAGGF 6
Db 24 EKAGGF 29

RESULT 8
US-10-366-125-19
; Sequence 19, Application US/10366125
; Publication No. US20030228259A1
; GENERAL INFORMATION:
; APPLICANT: Hellerstein, Marc
; TITLE OF INVENTION: MEASUREMENT OF BIOSYNTHESIS AND BREAKDOWN RATES OF
; TITLE OF INVENTION: BIOLOGICAL MOLECULES THAT ARE INACCESSIBLE OR NOT
; TITLE OF INVENTION: EASILY ACCESSIBLE TO DIRECT SAMPLING, NON-INVASIVELY,
; TITLE OF INVENTION: BY LABEL INCORPORATION INTO METABOLIC DERIVATIVES AND
; TITLE OF INVENTION: CATABOLIC PRODUCTS
; FILE REFERENCE: 416272003500
; CURRENT APPLICATION NUMBER: US/10/366,125
; CURRENT FILING DATE: 2003-02-12
; PRIOR APPLICATION NUMBER: US 60/356,008
; PRIOR FILING DATE: 2002-02-12
; NUMBER OF SEQ ID NOS: 28
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 19
; LENGTH: 262
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-366-125-19

Query Match      100.0%; Score 32; DB 15; Length 262;
Best Local Similarity 100.0%; Pred. No. 1.6e+02; Indels 0; Gaps 0;
Matches 6; Conservative 0; Mismatches 0;

QY 1 EKAGGF 6
Db 7 EKAGGF 12

RESULT 9
US-09-908-711-78
; Sequence 78, Application US/09908711
; Patent No. US20020045230A1
```

Mon Sep 20 11:05:42 2004

```
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PA128
; CURRENT APPLICATION NUMBER: US/09/908,711
; CURRENT FILING DATE: 2001-07-20
; PRIOR APPLICATION NUMBER: US01/01360
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,867
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01344
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,892
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01345
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,888
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01329
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,905
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01354
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,891
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01339
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,869
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01340
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,874
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01334
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,898
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01320
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,853
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01349
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,902
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01239
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,870
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01348
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,882
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01347
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,896
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01307
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,864
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01341
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,856
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01336
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,868
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01312
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 60/179,065
```

```
; PRIOR FILING DATE: 2000-01-31
; PRIOR APPLICATION NUMBER: 60/180,628
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: 60/209,467
; PRIOR FILING DATE: 2000-06-07
; NUMBER OF SEQ ID NOS: 167
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 78
; LENGTH: 309
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (2)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (4)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-908-711-78
```

```
Query Match 100.0%; Score 32; DB 9; Length 309;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

Qy 1 EKAGGF 6

Db 54 EKAGGF 59

## RESULT 10

US-10-437-963-162979

; Sequence 162979, Application US/10437963

; Publication No. US20040123343A1

; GENERAL INFORMATION:

; APPLICANT: La Rosa, Thomas J.

; APPLICANT: Kovalic, David K.

; APPLICANT: Zhou, Yihua

; APPLICANT: Cao, Yongwei

; APPLICANT: Wu, Wei

; APPLICANT: Boukharov, Andrey A.

; APPLICANT: Barbazuk, Brad

; APPLICANT: Li, Ping

; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With

; FILE REFERENCE: 38-21(53221)B

; CURRENT APPLICATION NUMBER: US/10/437,963

; CURRENT FILING DATE: 2003-05-14

; NUMBER OF SEQ ID NOS: 204966

; SEQ ID NO 162979

; LENGTH: 486

; TYPE: PRT

; ORGANISM: Oryza sativa

; FEATURE:

; OTHER INFORMATION: Clone ID: PAT\_MRT4530\_62017C.1.pep

US-10-437-963-162979

Query Match 100.0%; Score 32; DB 16; Length 486;

Best Local Similarity 100.0%; Pred. No. 3e+02;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 EKAGGF 6

Db 327 EKAGGF 332

## RESULT 11

US-10-058-124-21

; Sequence 21, Application US/10058124

; Publication No. US20030119058A1

; GENERAL INFORMATION:

; APPLICANT: Qvist, Per

; APPLICANT: Bonde, Martin

; TITLE OF INVENTION: A Method for Assaying Collagen Fragments



in Body Fluids, A Test Kit and Means for Carrying Out the  
Method and Use of the Method to Diagnose the Presence of  
Disorders Associated with the Metabolism of

## NUMBER OF SEQUENCES: 21

## CORRESPONDENCE ADDRESS:

ADDRESSEE: Darby & Darby PC  
STREET: 805 Third Avenue  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10022

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25

## CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/058,124

FILING DATE: 29-Jan-2002

CLASSIFICATION: <Unknown>

## PRIOR APPLICATION DATA:

APPLICATION NUMBER: 09/570,573

FILING DATE: 2002-MAY-12

APPLICATION NUMBER: 08/187,319

## ATTORNEY/AGENT INFORMATION:

NAME: Gogoris, Adda C

REGISTRATION NUMBER: 29,714

REFERENCE/DOCKET NUMBER: 4305/08701

## TELECOMMUNICATION INFORMATION:

TELEPHONE: 212-527-7700

TELEFAX: 212-753-6237

## TELEX: 236687

## INFORMATION FOR SEQ ID NO: 21:

## SEQUENCE CHARACTERISTICS:

LENGTH: 1078 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

ORIGINAL SOURCE:

ORGANISM: Homo sapiens

IMMEDIATE SOURCE:

CLONE: COLLAGEN ALPHA 1 (III)

SEQUENCE DESCRIPTION: SEQ ID NO: 21:

US-10-058-124-21

Query Match 100.0%; Score 32; DB 14; Length 1078;

Best Local Similarity 100.0%; Pred. No. 6.6e+02;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 EKAGGF 6

Db 1064 EKAGGF 1069

## RESULT 12

US-09-918-715-226

Sequence 226, Application US/09918715

Publication No. US20030017157A1

## GENERAL INFORMATION:

APPLICANT: Brad St. Croix

APPLICANT: Bert Vogelstein

APPLICANT: Kenneth Kinzler

TITLE OF INVENTION: ENDOTHELIAL CELL EXPRESSION PATTERNS

FILE REFERENCE: 1107.00134

CURRENT APPLICATION NUMBER: US/09/918,715

CURRENT FILING DATE: 2001-08-01

PRIOR APPLICATION NUMBER: 60/222,599

PRIOR FILING DATE: 2000-08-02

PRIOR APPLICATION NUMBER: 60/224,360

PRIOR FILING DATE: 2000-08-11

PRIOR APPLICATION NUMBER: 60/282,850

PRIOR FILING DATE: 2000-04-11

; NUMBER OF SEQ ID NOS: 358

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 226

; LENGTH: 1466

; TYPE: PRT

; ORGANISM: Homo sapiens

US-09-918-715-226

Query Match 100.0%; Score 32; DB 12; Length 1466;

Best Local Similarity 100.0%; Pred. No. 9e+02;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 EKAGGF 6

Db 1211 EKAGGF 1216

## RESULT 13

US-10-257-021-72

Sequence 72, Application US/10257021

Publication No. US20030211498A1

## GENERAL INFORMATION:

APPLICANT: Morin, Patrice J.

APPLICANT: Sherman-Baust, Cheryl A.

APPLICANT: Pizer, Ellen S.

APPLICANT: Hough, Colleen D.

TITLE OF INVENTION: TUMOR MARKERS IN OVARIAN CANCER

FILE REFERENCE: 14014.0369U2

CURRENT APPLICATION NUMBER: US/10/257,021

CURRENT FILING DATE: 2002-10-03

PRIOR APPLICATION NUMBER: PCT/US01/10947

PRIOR FILING DATE: 2001-04-03

PRIOR APPLICATION NUMBER: 60/194,336

PRIOR FILING DATE: 2000-04-03

NUMBER OF SEQ ID NOS: 147

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 72

LENGTH: 1466

TYPE: PRT

ORGANISM: Homo sapiens

US-10-257-021-72

Query Match 100.0%; Score 32; DB 12; Length 1466;

Best Local Similarity 100.0%; Pred. No. 9e+02;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 EKAGGF 6

Db 1211 EKAGGF 1216

## RESULT 14

US-10-177-293-68

Sequence 68, Application US/10177293

Publication No. US20030124128A1

## GENERAL INFORMATION:

APPLICANT: Lillie, James

APPLICANT: Glatt, Karen

APPLICANT: Zhao, Xumei

APPLICANT: Ganavarpu, Manjula

APPLICANT: Kamatkar, Shubhangi

APPLICANT: Mertens, Maureen

APPLICANT: Myer, Vic

APPLICANT: Wang, Youzhen

APPLICANT: Xu, Yongyao

APPLICANT: Hoersch, Sebastian

APPLICANT: Monahan, John

APPLICANT: Meyers, Rachel E.

APPLICANT: Bast Jr., Robert C.

APPLICANT: Hortobagyi, Gabriel N.

APPLICANT: Pusztai, Lajos

APPLICANT: Meric, Funda

APPLICANT: Sahin, Aysegul

Mon Sep 20 11:05:42 2004

```

; APPLICANT: Mills, Gordon B.
; TITLE OF INVENTION: COMPOSITIONS, KITS, AND METHODS FOR IDENTIFICATION, ASSESSMENT,
; TITLE OF INVENTION: PREVENTION, AND THERAPY OF BREAST CANCER
; FILE REFERENCE: MRI-038
; CURRENT APPLICATION NUMBER: US/10/177,293
; CURRENT FILING DATE: 2002-06-21
; PRIOR APPLICATION NUMBER: US 60/299,887
; PRIOR FILING DATE: 2001-06-21
; PRIOR APPLICATION NUMBER: US 60/301,572
; PRIOR FILING DATE: 2001-06-27
; PRIOR APPLICATION NUMBER: US 60/306,501
; PRIOR FILING DATE: 2001-07-18
; PRIOR APPLICATION NUMBER: US 60/325,002
; PRIOR FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: US 60/362,585
; PRIOR FILING DATE: 2002-03-05
; PRIOR APPLICATION NUMBER: US 60/xxx,xxx
; PRIOR FILING DATE: 2002-05-14
; NUMBER OF SEQ ID NOS: 506
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 68
; LENGTH: 1466
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-177-293-68

Query Match 100.0%; Score 32; DB 14; Length 1466;
Best Local Similarity 100.0%; Pred. No. 9e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EKAGGF 6
Db 1211 EKAGGF 1216

RESULT 15
US-10-301-822-33
; Sequence 33, Application US/10301822
; Publication No. US20030148410A1
; GENERAL INFORMATION:
; APPLICANT: Millennium Pharmaceuticals, Inc.
; APPLICANT: Berger, Allison
; APPLICANT: Guillemette, Tracy L.
; APPLICANT: Kamatkar, Shubhangi
; APPLICANT: Schlegel, Robert
; APPLICANT: Monahan, John E.
; APPLICANT: Thibodeau, Stephen N.
; APPLICANT: Burgart, Lawrence J.
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND
; TITLE OF INVENTION: METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF COLON CANCER
; FILE REFERENCE: MPM01-029P2RNM
; CURRENT APPLICATION NUMBER: US/10/301,822
; CURRENT FILING DATE: 2002-11-21
; PRIOR APPLICATION NUMBER: US 60/339,971
; PRIOR FILING DATE: 2001-12-10
; PRIOR APPLICATION NUMBER: US 60/361,978
; PRIOR FILING DATE: 2002-03-05
; PRIOR APPLICATION NUMBER: US 60/381,988
; PRIOR FILING DATE: 2002-05-20
; NUMBER OF SEQ ID NOS: 228
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 33
; LENGTH: 1466
; TYPE: PRT
; ORGANISM: Homo Sapiens
US-10-301-822-33

Query Match 100.0%; Score 32; DB 14; Length 1466;
Best Local Similarity 100.0%; Pred. No. 9e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EKAGGF 6
```

Db 1211 EKAGGF 1216

Search completed: September 18, 2004, 04:56:11

Job time : 29.4706 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 18, 2004, 04:02:05 ; Search time 7.29412 Seconds  
(without alignments)  
79.125 Million cell updates/sec

Title: US-10-615-959-39  
Perfect score: 32  
Sequence: 1 EKAGGF 6

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_78:.\*  
1: pir1.\*  
2: pir2.\*  
3: pir3.\*  
4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	32	100.0	97	2 B71719	ribosomal protein
2	32	100.0	97	2 H97716	50S ribosomal prot
3	32	100.0	1289	2 F72308	hypothetical prote
4	32	100.0	1466	1 CGHU7L	collagen alpha 1(I
5	29	90.6	187	2 H87376	hypothetical prote
6	29	90.6	300	2 A55804	phosphoprotein pho
7	29	90.6	382	2 S55482	probable NADPH deh
8	29	90.6	416	1 KIHOG	phosphoglycerate k
9	29	90.6	417	1 KIHUG	phosphoglycerate k
10	29	90.6	417	2 A25567	phosphoglycerate k
11	29	90.6	417	2 I48074	phosphoglycerate k
12	29	90.6	417	2 A33792	phosphoglycerate k
13	29	90.6	421	2 B84702	phosphoglycerate k
14	29	90.6	457	2 B87269	hypothetical prote
15	29	90.6	522	2 T28687	hypothetical prote
16	29	90.6	522	2 S62941	hypothetical prote
17	29	90.6	631	2 C64034	hypothetical prote
18	29	90.6	636	2 S41067	hypothetical prote
19	29	90.6	1028	2 B24785	collagen alpha 1(I
20	29	90.6	1075	2 G85360	hypothetical prote
21	29	90.6	1440	2 C84639	puative protein [i
22	29	90.6	1464	2 S59856	hypothetical prote
23	28	87.5	127	2 E81134	collagen alpha 1(I
24	28	87.5	127	2 B81885	hypothetical prote
25	28	87.5	240	2 AF2443	probable membrane
26	28	87.5	294	2 E83022	hypothetical prote
27	28	87.5	306	2 D69298	conserved glycosyl
28	28	87.5	319	2 A70396	hypothetical prote
29	28	87.5	355	2 F69298	conserved hypothet

30	28	87.5	355	2 C95843	probable DNA ligas
31	28	87.5	371	2 G87644	hypothetical prote
32	28	87.5	403	1 F69209	conserved hypothet
33	28	87.5	585	2 B69218	conserved hypothet
34	28	87.5	644	2 T02366	hypothetical prote
35	28	87.5	655	2 T51792	hypothetical prote
36	28	87.5	1559	2 T30535	hypothetical prote
37	28	87.5	1571	2 AC1647	calcium channel al
38	28	87.5	1852	2 A37860	Lactobacillus phag
39	27	84.4	115	2 S24663	calcium channel pr
40	27	84.4	120	2 AC3550	Ig heavy chain v r
41	27	84.4	124	2 E90326	NAD(P) transhydrog
42	27	84.4	140	2 D72480	hypothetical prote
43	27	84.4	147	2 S76645	phosphohistidine p
44	27	84.4	223	2 E72632	hypothetical prote
45	27	84.4	231	2 AB1038	hypothetical prote

## ALIGNMENTS

RESULT 1  
B71719  
ribosomal protein L28 - Rickettsia prowazekii  
C:Species: Rickettsia prowazekii  
C:Date: 21-Nov-1998 #sequence\_revision 21-Nov-1998 #text\_change 03-Nov-2000  
R:Accession: B71719  
R:Andersson, S.G.E.; Zomorodipour, A.; Andersson, J.O.; Sicheritz-Ponten, T.; Alsmark, U.  
Nature 396, 133-140, 1998  
A:Title: The genome sequence of Rickettsia prowazekii and the origin of mitochondria.  
A:Reference number: A71630; MUID:99039499; PMID:9823893  
A:Accession: B71719  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-97 <AND>  
A:Cross-references: GB:AJ235270; GB:AJ235269; NID:G3860572; PIDN:CAA14569.1; PID:ei134241  
A:Experimental source: strain Madrid E  
C:Genetics:  
A:Gene: rpmb; RP099

Query Match 100.0%; Score 32; DB 2; Length 97;  
Best Local Similarity 100.0%; Pred. No. 8.9;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 EKAGGF 6  
|||  
Db 60 EKAGGF 65

## RESULT 2

H97716  
50S ribosomal protein L28 [imported] - Rickettsia conorii (strain Malish 7)  
C:Species: Rickettsia conorii  
C:Date: 30-Sep-2001 #sequence\_revision 30-Sep-2001 #text\_change 30-Sep-2001  
C:Accession: H97716  
R:Ogata, H.; Audic, S.; Renesto-Audiffren, P.; Fournier, P.E.; Barbe, V.; Samson, D.; Rou  
Science 293, 2093-2098, 2001  
A:Title: Mechanisms of Evolution in Rickettsia conorii and Rickettsia prowazekii.  
A:Reference number: A97700; MUID:2142074; PMID:11557893  
A:Accession: H97716  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-97 <KUR>  
A:Cross-references: GB:AE006914; PIDN:AAL02674.1; PID:gl5619179; GSPDB:GN00173  
C:Genetics:  
A:Gene: rpmb

Query Match 100.0%; Score 32; DB 2; Length 97;  
Best Local Similarity 100.0%; Pred. No. 8.9;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 EKAGGF 6  
|||

Db 60 EKAGGF 65

# RESULT 3

W72308  
 Hypothetical protein - Thermotoga maritima (strain MSB8)  
 C:Species: Thermotoga maritima  
 C:Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 21-Jul-2000  
 C:Accession: F72308  
 R:Nelson, K.B.; Clayton, R.A.; Gill, S.R.; Swinn, M.L.; Dodson, R.J.; Haft, D.H.; Hickey  
 Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.;  
 C.M.  
 Nature 399, 323-329, 1999  
 A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome seq  
 A:Reference number: A72200; PMID:99287316; PMID:10360571  
 A:Accession: F72308  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-1289 <ARN>  
 A:Cross-references: GB:AE001761; GB:AE000512; NID:94981529; PIDN:AAD36071.1; PID:9498153  
 A:Experimental source: strain MSB8  
 C:Genetics:  
 A:Gene: TM0992

Query Match	100.0%	Score 32;	DB 2;	Length 1289;
Best Local Similarity	100.0%	Pred. NO. 98;		
Matches	6;	Conservative 0;	Mismatches 0;	Indels 0;
				Gaps 0;

QY 1 EKAGGF 6

Db 942 EKAGGF 947

# RESULT 4

CGH07L  
 collagen alpha 1(III) chain precursor - human  
 N:Alternate names: procollagen alpha 1(III) chain  
 C:Species: Homo sapiens (man)  
 C:Date: 24-Apr-1984 #sequence\_revision 01-Sep-1995 #text\_change 21-Jul-2000  
 C:Accession: S05272; P0011; S01726; S04887; A90399; A94562; I51868; S59511; A90  
 R:Prockop, D.J.  
 submitted to the EMBL Data Library, February 1989  
 A:Reference number: S05272  
 A:Accession: S05272  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-1240, 'Y', 1242-1466 <PRC>  
 A:Cross-references: EMBL:X14420; NID:G30057; PIDN:CAA32583.1; PID:G30058  
 R:Ala-Kokko, L.; Kontusaari, S.; Baldwin, C.T.; Kulvanniemi, H.; Prockop, D.J.  
 Biochem. J. 260, 509-516, 1989  
 A:Title: Structure of cDNA clones coding for the entire prepro-alpha(III) chain of huma  
 erences.  
 A:Reference number: S04642; PMID:89350838; PMID:2764886  
 A:Accession: S04642  
 A:Molecule type: mRNA  
 A:Residues: 1-1196 <ALA>  
 A:Cross-references: EMBL:X14420; NID:G30057; PIDN:CAA32583.1; PID:G30058  
 A:Note: the complete sequence is not shown  
 R:Benson-Chanda, V.; Su, M.W.; Weil, D.; Chu, M.L.; Ramirez, F.  
 Gene 78, 255-265, 1989  
 A:Title: Cloning and analysis of the 5' portion of the human type-III procollagen gene  
 A:Reference number: P0011; PMID:89378752; PMID:2777083  
 A:Accession: P0011  
 A:Molecule type: DNA  
 A:Residues: 1-176 <BEN>  
 A:Cross-references: NID:G180813; PIDN:AAA52040.1; PID:G180814  
 R:Toman, P.D.; Ricca, G.A.; de Crombrughe, B.  
 Nucleic Acids Res. 16, 7201, 1988  
 A:Title: Nucleotide sequence of a cDNA coding for the amino-terminal region of human pre  
 A:Reference number: S01726; PMID:88303360; PMID:3405773  
 A:Accession: S01726  
 A:Molecule type: mRNA  
 A:Residues: 1-170 <TM>

A:Cross-references: EMBL:X07240; NID:G30060; PIDN:CAA30229.1; PID:G30061  
 A:Note: the authors translated the codon CAG for residue 154 as His  
 R:Janeczko, R.A.; Ramirez, F.  
 Nucleic Acids Res. 17, 6742, 1989  
 A:Title: Nucleotide and amino acid sequences of the entire human alpha-1 (III) collagen.  
 A:Reference number: S04887; PMID:89386015; PMID:2780304  
 A:Accession: S04887

A:Molecule type: mRNA  
 A:Residues: 149-163, 'G', 164-240, 'D', 242-471, 'D', 473-487, 'L', 489, 'S', 491-613, 'Y', 615-634, '  
 A:Cross-references: EMBL:X15332; NID:G939545; PIDN:CAA33387.1; PID:G9390045  
 A:Note: the authors' translation of residues 905-932 is inconsistent with the nucleotide  
 R:Seyer, J.M.; Kang, A.H.  
 Biochemistry 16, 1158-1164, 1977  
 A:Title: Covalent structure of collagen: amino acid sequence of cyanogen bromide peptides  
 A:Reference number: A90399; PMID:77134724; PMID:557335  
 A:Accession: A90399  
 A:Molecule type: protein  
 A:Residues: 'V', 169-225, 229-232, 'P', 234-292, 'D', 294-398 <SEY1>  
 A:Experimental source: liver  
 A:Note: sequence corrected by A94562; attachment of 2-O-alpha-D-glucosyl-O-beta-D-galact  
 R:Seyer, J.M.  
 submitted to the Atlas, December 1977  
 A:Reference number: A94562  
 A:Accession: A94562  
 A:Molecule type: protein  
 A:Residues: 'V', 169-225, 229-277, 'A', 279-292, 'D', 294, 'S', 296-398 <SEY2>  
 A:Experimental source: liver  
 A:Note: author submitted corrections to A90399  
 R:Miliewicz, D.M.; Witz, A.M.; Smith, A.C.; Manchester, D.K.; Waldstein, G.; Byers, P.H.  
 Am. J. Hum. Genet. 53, 62-70, 1993  
 A:Title: Parental somatic and germ-line mosaicism for a multiexon deletion with unusual  
 Espring.  
 A:Reference number: I51868; PMID:93304430; PMID:8317500  
 A:Accession: I51868  
 A:Status: preliminary; translated from GB/EMBL/DBDJ  
 A:Molecule type: DNA  
 A:Residues: 186-194 <MIL>  
 A:Cross-references: GB:S62925; NID:G386425; PIDN:AAD13937.1; PID:G4261637  
 R:Chiodo, A.A.; Sillence, D.O.; Cole, W.G.; Bateman, J.F.  
 Biochem. J. 311, 939-943, 1995  
 A:Title: Abnormal type III collagen produced by an exon-17-skipping mutation of the COL3A  
 A:Reference number: S59511; PMID:96067614; PMID:7487954  
 A:Accession: S59511  
 A:Molecule type: mRNA  
 A:Residues: 302-423 <CHI>  
 A:Cross-references: GB:S79877; NID:G1195576; PIDN:AAB35615.1; PID:G1195577  
 R:Seyer, J.M.; Kang, A.H.  
 Biochemistry 17, 3404-3411, 1978  
 A:Title: Covalent structure of collagen: amino acid sequence of five consecutive CNBr pep  
 A:Reference number: A90414; PMID:79000343; PMID:687591  
 A:Accession: A90414  
 A:Molecule type: protein  
 A:Residues: 399-675, 'N', 677-727 <SEY3>  
 A:Experimental source: liver  
 R:Lee, B.; Vitale, E.; Superti-Furga, A.; Steinmann, B.; Ramirez, F.  
 J. Biol. Chem. 266, 5256-5259, 1991  
 A:Title: G to T transversion at position +5 of a splice donor site causes skipping of the  
 A:Reference number: I55349; PMID:91161621; PMID:1672129  
 A:Accession: I55349  
 A:Status: translated from GB/EMBL/DBDJ  
 A:Molecule type: DNA  
 A:Residues: 537-605 <LEE>  
 A:Cross-references: GB:M59312; NID:G180815; PIDN:AAA52041.1; PID:G180816  
 R:Seyer, J.M.; Mainardi, C.; Kang, A.H.  
 Biochemistry 19, 1583-1589, 1980  
 A:Title: Covalent structure of collagen: amino acid sequence of alpha1 (III)-CB5 from tom  
 A:Reference number: A90438; PMID:80198282; PMID:6246925  
 A:Accession: A90438  
 A:Molecule type: protein  
 A:Residues: 728-895, 'A', 897-964 <SEY4>  
 A:Experimental source: liver  
 R:Cole, W.G.; Chiodo, A.A.; Lamande, S.R.; Janeczko, R.; Ramirez, F.; Dahl, H.H.M.; Chan  
 J. Biol. Chem. 265, 17070-17077, 1990

A:Title: A base substitution at a splice site in the COL3A1 gene causes exon skipping and  
 A:Reference number: A38303; MUID:91009133; PMID:2145268  
 A:Accession: A38303  
 A:Molecule type: mRNA  
 A:Residues: 861-1015 <COL>  
 A:Cross-references: GB:J05617; GB:M55603; GB:M59227; NID:G180878; PIDN:AA559383.1; PID:9  
 A:Note: A mutant sequence with 942-977 spliced out from a patient with Ehlers-Danlos syn  
 R:Mankoo, B.S.; Dalgleish, R.  
 Nucleic Acids Res. 16, 2337, 1988  
 A:Title: Human pro alpha1(III) collagen: cDNA sequence for the 3' end.  
 A:Reference number: S02119; MUID:88189827; PMID:3357782  
 A:Accession: S02119  
 A>Status: translation not shown  
 A:Molecule type: mRNA  
 A:Residues: 950-1018, 'V', 1020-1193, 'S', 1195-1466 <MAN>  
 A:Cross-references: EMBL:X06700; NID:G30053; PIDN:CAA29886.1; PID:G30054  
 R:Seyer, J.M.; Kang, A.H.  
 Biochemistry 20, 2621-2627, 1981  
 A:Title: Covalent structure of collagen: amino acid sequence of alpha1 (III)-CB9 from ty  
 A:Reference number: A90446; MUID:81208139; PMID:7016180  
 A:Accession: A90446  
 A:Molecule type: protein  
 A:Residues: 965-979, 'A', 981-984, 'PS', 987, 'QN', 990-1096, 'P', 1098-1152, 'AT', 1155, 'S', 1157-  
 R:Loidl, H.R.; Brinker, J.M.; May, M.; Pihlajaniemi, T.; Morrow, S.; Rosenbloom, J.; Mye  
 Nucleic Acids Res. 12, 9383-9394, 1984  
 A:Title: Molecular cloning and carboxyl-propeptide analysis of human type III procollage  
 A:Reference number: A93551; MUID:85087944; PMID:6096827  
 A:Accession: A93551  
 A:Molecule type: mRNA  
 A:Residues: 1065-1155, 'P', 1157-1466 <LOI>  
 A:Cross-references: EMBL:X01655; EMBL:X01742; NID:G29584; PIDN:CAA25821.1  
 R:Miskulin, M.; Dalgleish, R.; Kluge-Beckerman, B.; Renard, S.I.; Tolstoshev, P.; Brant  
 Biochemistry 25, 1408-1413, 1986  
 A:Title: Human type III collagen gene expression is coordinately modulated with the type  
 A:Reference number: 152393; MUID:86187804; PMID:3754462  
 A:Accession: 152393  
 A>Status: translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1161-1200 <MIS>  
 A:Cross-references: GB:M13146; NID:G180415; PIDN:AA52003.1; PID:G180416  
 R:Emanuel, B.S.; Cannizzaro, L.A.; Seyer, J.M.; Myers, J.C.  
 Proc. Natl. Acad. Sci. U.S.A. 82, 3385-3389, 1985  
 A:Title: Human alpha 1(III) and alpha 2(V) procollagen genes are located on the long arm  
 A:Reference number: 159025; MUID:85216505; PMID:3858826  
 A:Accession: 179359  
 A>Status: translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1165-1196 <EMA>  
 A:Cross-references: GB:M11134; NID:G180417; PIDN:AA52004.1; PID:G180418  
 R:Chu, M.L.; Weil, D.; de Wet, W.; Bernard, M.; Sippola, M.; Ramirez, F.  
 J. Biol. Chem. 260, 4357-4363, 1985  
 A:Title: Isolation of cDNA and genomic clones encoding human pro-alpha1(III) collagen. F  
 A:Reference number: A92516; MUID:85157600; PMID:2579949  
 A:Accession: A92516  
 A:Molecule type: DNA  
 A:Residues: 1176-1240, 'V', 1242-1356, 'P', 1358-1466 <CHU>  
 A:Cross-references: GB:M10635; GB:M10793; GB:M10794; GB:M10795; GB:M10796; GB:M10797; GB  
 A:Experimental source: liver  
 A:Note: the authors translated the codon TTC for residue 1057 as Tyr; the codons given f  
 ation  
 C:Comment: Prolines and lysines at the third position of the tripeptide repeating unit  
 3-hydroxylated. About 15% of the lysines are 5-hydroxylated and some are subsequently C  
 C:Genetics:  
 A:Gene: GDB:COL3A1  
 A:Map position: 2q31-2q31  
 A:Introns: 27/1; 94/3; 111/3; 149/3; 176/3; 554/3; 587/3; 1175/3; 1275/1; 1337/3; 1418/3  
 A:Note: the list of introns is incomplete; defects in this gene can result in Ehlers-Dan  
 C:Complex: type III collagen is a homotrimer of monomers initially linked by disulfide b  
 er of their length, is formed with desmosine cross-links made from lysine and allylsine  
 C:Function:  
 A:Description: structural component of extracellular fibrous polymer that maintains inte

C:Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;  
 C:Keywords: coiled coil; Ehlers-Danlos syndrome; extracellular matrix; glycoprotein; hydn  
 F:1-23/Domain: signal sequence #status predicted <SIG>  
 F:24-153/Domain: amino-terminal propeptide #status predicted <PRO>  
 F:31-91/Domain: von Willebrand factor type C repeat homology <VWC>  
 F:154-1221/Product: collagen alpha 1(III) chain #status predicted <MAT>  
 F:154-167/Region: amino-terminal nonhelical telopeptide  
 F:168-1196/Region: helical  
 F:1091-1093/Region: cell attachment (R-G-D) motif  
 F:1197-1221/Region: carboxyl-terminal nonhelical telopeptide  
 F:1222-1466/Domain: carboxyl-terminal propeptide #status predicted <CDP>  
 F:1238-1466/Domain: fibrillar collagen carboxyl-terminal homology <FCC>  
 F:24/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status predicted  
 F:153-154/Cleavage site: Pro-Gln (procollagen N-endopeptidase) #status predicted  
 F:154/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status predicted  
 F:161.1212/Modified site: allylsine (Lys) #status predicted  
 F:263.284, 860, 977, 1106/Modified site: 5-hydroxylysine (Lys) #status experimental  
 F:263/Binding site: carbohydrate (Lys) (covalent) #status experimental  
 F:584, 1034/Modified site: 5-hydroxylysine (Lys) (partial) #status experimental  
 F:948-949/Cleavage site: Gly-Ile (collagenase) #status experimental  
 F:1106/Binding site: carbohydrate (Lys) (covalent) #status predicted  
 Query Match 100.0%; Score 32; DB 1; Length 1466;  
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 1 EKAGGF 6  
 Db 1211 EKAGGF 1216  
 RESULT 5  
 H87376  
 Hypothetical protein CC1028 [imported] - Caulobacter crescentus  
 C:Species: Caulobacter crescentus  
 C:Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 20-Apr-2001  
 C:Accession: H87376  
 R:Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.I  
 B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolona  
 n, J.; Rasmussen, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.  
 Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001  
 A:Title: Complete Genome Sequence of Caulobacter crescentus.  
 A:Reference number: AB7249; MUID:21173698; PMID:11259647  
 A:Accession: H87376  
 A>Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-187 <STO>  
 A:Cross-references: GB:AF005673; NID:G13422320; PIDN:AAK23012.1; GSPDB:GN00148  
 C:Genetics:  
 Query Match 90.6%; Score 29; DB 2; Length 187;  
 Best Local Similarity 83.3%; Pred. No. 73;  
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 Qy 1 EKAGGF 6  
 Db 126 ERAGGF 131  
 RESULT 6  
 A55804  
 phosphoprotein phosphatase (EC 3.1.3.16) 2c, membrane-bound - Paramesidium tetraurelia  
 C:Species: Paramesidium tetraurelia  
 C:Date: 23-Mar-1995 #sequence\_revision 05-Apr-1995 #text\_change 07-Dec-1999  
 C:Accession: A55804  
 R:Klump, S.; Hanke, C.; Donella-Deana, A.; Beyer, A.; Kellner, R.; Pinna, L.A.; Schultze,  
 J. Biol. Chem. 269, 32774-32780, 1994  
 A:Title: A membrane-bound protein phosphatase type 2C from Paramesidium tetraurelia. Purifi  
 A:Reference number: A55804; MUID:95105156; PMID:7806499  
 A:Accession: A55804  
 A>Status: preliminary  
 A:Molecule type: mRNA

Mon Sep 20 11:05:42 2004

us-10-615-959-39.rpr

C;Comment: Residues thought to be involved in ADP-ATP binding are Glu-343 to ribose and 139-341.  
C;Superfamily: phosphoglycerate kinase  
C;Keywords: acetylated amino end; ATP; gluconogenesis; glycolysis; phosphotransferase  
F;1/Modified site: acetylated amino end (Ser) #status experimental  
F;219,343/Binding site: ATP (Iys, Glu) #status experimental

Query Match 90.6%; Score 29; DB 1; Length 416;  
Best Local Similarity 83.3%; Pred. No. 1.5e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 EKAGGF 6  
Db 182 QKAGGF 187

RESULT 9

KIHUG  
phosphoglycerate kinase (EC 2.7.2.3) - human  
N;Alternate names: primer recognition protein 2  
C;Species: Homo sapiens (man)  
C;Date: 14-Nov-1983 #sequence\_revision 23-Aug-1996 #text\_change 24-Sep-1999  
C;Accession: I59050; A00669; A92279; A35739; A22426; I59539  
R;Michelson, A.M.; Blake, C.C.; Evans, S.T.; Orkin, S.H.  
Proc. Natl. Acad. Sci. U.S.A. 82, 6965-6969, 1985  
A;Title: Structure of the human phosphoglycerate kinase gene and the intron-mediated evolution  
A;Reference number: I59050; MUID:86016816; PMID:2995995  
A;Accession: I59050  
A;Status: translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-417 <RE2>  
A;Cross-references: GB:M11968; NID:gl89920; PIDN:AAA60079.1; PID:g387021  
R;Michelson, A.M.; Markham, A.F.; Orkin, S.H.  
Proc. Natl. Acad. Sci. U.S.A. 80, 472-476, 1983  
A;Title: Isolation and DNA sequence of a full-length cDNA clone for human X chromosome-  
A;Reference number: A93962; MUID:83169680; PMID:6188151  
A;Accession: A00669  
A;Molecule type: mRNA  
A;Residues: 1-417 <MIC>  
A;Cross-references: GB:L00160; GB:J00288; NID:gl89904; PIDN:AAA60078.1; PID:g387020  
R;Experimental source: liver  
R;Huang, I.Y.; Welch, C.D.; Yoshida, A.  
J. Biol. Chem. 255, 6412-6420, 1980  
A;Title: Complete amino acid sequence of human phosphoglycerate kinase. Cyanogen bromide  
A;Reference number: A92279; MUID:80227775; PMID:7391027  
A;Accession: A92279  
A;Molecule type: protein  
A;Residues: 2-39,'K',40-52,'D',54-109,'D',111-275,'D',277-299,'E',301-336,'D',338-385,'Q'  
A;Experimental source: erythrocytes  
R;Jindal, H.K.; Vishwanatha, J.K.  
J. Biol. Chem. 265, 6540-6543, 1990  
A;Title: Functional identity of a primer recognition protein as phosphoglycerate kinase.  
A;Reference number: A35739; MUID:90216667; PMID:2324090  
A;Accession: A35739  
A;Molecule type: protein  
A;Residues: 2-6;42-48;142-151;247-264;298-322;324-330;383-406 <JIN>  
R;Singer-Sam, J.; Keith, D.H.; Tani, K.; Simmer, R.L.; Shively, L.; Lindsay, S.; Yoshida  
Gene 32, 409-417, 1984  
A;Title: Sequence of the promoter region of the gene for human X-linked 3-phosphoglycerate  
A;Reference number: A22426; MUID:85155507; PMID:6099325  
A;Accession: A22426  
A;Molecule type: DNA  
A;Residues: 1-21 <SIN>  
R;Pfeifer, G.P.; Steigerwald, S.D.; Mueller, P.R.; Wold, B.; Riggs, A.D.  
Science 246, 810-813, 1989  
A;Title: Genomic sequencing and methylation analysis by ligation mediated PCR.  
A;Reference number: I59539; MUID:90049205; PMID:2814502  
A;Accession: I59539  
A;Status: translation not shown; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-14 <RES>  
A;Cross-references: GB:M34017; NID:gl89993; PIDN:AAA60103.1; PID:gl89994  
C;Comment: The structure consists of two discrete, globular domains that are joined by a

A;Residues: 1-300 <KLU>  
A;Cross-references: GB:Z36985; NID:g537421; PID:e1192609; PID:g2654382  
C;Species: Schizosaccharomyces pombe  
C;Date: 01-Aug-1995 #sequence\_revision 01-Sep-1995 #text\_change 10-Dec-1999  
C;Accession: S55482; T38968  
R;Connor, R.; Churcher, C.M.  
submitted to the EMBL Data Library, May 1995  
A;Reference number: S55479  
A;Accession: S55482  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-382 <CON>  
A;Cross-references: EMBL:Z49811; NID:g854599; PIDN:CAA89954.1; PID:g854603  
R;Connor, R.; Churcher, C.M.; Barrell, B.G.; Rajandream, M.A.; Walsh, S.V.  
submitted to the EMBL Data Library, May 1995  
A;Reference number: Z21821  
A;Accession: T38968  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-382 <CO2>  
A;Cross-references: EMBL:Z49811; PIDN:CAA89954.1; GSPDB:GN00066; SPDB:SPAC5H10.04  
A;Experimental source: strain 972h-; cosmid c5H10  
C;Genetics:  
A;Gene: SPDB:SPAC5H10.04  
A;Map Position: 1  
C;Superfamily: NADPH dehydrogenase chain OYE2

Query Match 90.6%; Score 29; DB 2; Length 300;  
Best Local Similarity 83.3%; Pred. No. 1.1e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 EKAGGF 6  
Db 175 ERAGGF 180

RESULT 7

probable NADPH dehydrogenase - fission yeast (Schizosaccharomyces pombe)  
C;Species: Schizosaccharomyces pombe  
C;Date: 01-Aug-1995 #sequence\_revision 01-Sep-1995 #text\_change 10-Dec-1999  
C;Accession: S55482; T38968  
R;Connor, R.; Churcher, C.M.  
submitted to the EMBL Data Library, May 1995  
A;Reference number: S55479  
A;Accession: S55482  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-382 <CON>  
A;Cross-references: EMBL:Z49811; NID:g854599; PIDN:CAA89954.1; PID:g854603  
R;Connor, R.; Churcher, C.M.; Barrell, B.G.; Rajandream, M.A.; Walsh, S.V.  
submitted to the EMBL Data Library, May 1995  
A;Reference number: Z21821  
A;Accession: T38968  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-382 <CO2>  
A;Cross-references: EMBL:Z49811; PIDN:CAA89954.1; GSPDB:GN00066; SPDB:SPAC5H10.04  
A;Experimental source: strain 972h-; cosmid c5H10  
C;Genetics:  
A;Gene: SPDB:SPAC5H10.04  
A;Map Position: 1  
C;Superfamily: NADPH dehydrogenase chain OYE2

Query Match 90.6%; Score 29; DB 2; Length 382;  
Best Local Similarity 83.3%; Pred. No. 1.4e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 EKAGGF 6  
Db 66 EKGGGF 71

RESULT 8

KIHUG  
phosphoglycerate kinase (EC 2.7.2.3) - horse  
C;Species: Equus caballus (domestic horse)  
C;Date: 31-Mar-1992 #sequence\_revision 31-Mar-1992 #text\_change 28-Feb-1997  
C;Accession: A92292; A00669  
R;Marrett, M.  
J. Biol. Chem. 256, 10293-10305, 1981  
A;Title: Primary structure of 3-phosphoglycerate kinase from horse muscle. II. Amino acid  
complete sequence of the enzyme.  
A;Reference number: A92292; MUID:82030789; PMID:7287713  
A;Accession: A92292  
A;Molecule type: protein  
A;Residues: 1-416 <MER>  
R;Banks, R.D.; Blake, C.C.F.; Evans, P.R.; Hase, R.; Rice, D.W.; Hardy, G.W.; Merrett,  
Nature 279, 773-777, 1979  
A;Title: Sequence, structure and activity of phosphoglycerate kinase: a possible hinge-  
A;Reference number: A93209; MUID:79199779; PMID:450128  
C;Contents: annotation; X-ray crystallography, 2.5 angstroms; muscle  
C;Comment: The structure consists of two discrete, globular domains that are joined by a  
405-416 form a helix associated with the amino-terminal domain.

406-417 form a helix associated with the amino-terminal domain.  
C:Comment: Residues thought to be involved in ADP-ATP binding are Glu-344 to ribose and 40-342.

C:Genetics:  
A:Gene: GDB:PGK1  
A:Cross-references: GDB:120282; OMIM:311800  
A:Map position: Xq13.3-Xq13.3  
A:Introns: 22/2; 39/2; 91/2; 139/3; 174/2; 214/2; 252/3; 312/3; 372/1; 405/1  
C:Superfamily: phosphoglycerate kinase  
C:Keywords: ATP; blocked amino end; gluconeogenesis; glycolysis; phosphotransferase  
F:2/Modified site: blocked amino end (Ser) (in mature form) (probably acetylated) #status  
F:220,344/Binding site: ATP (Lys, Glu) #status Predicted

Query Match 90.6%; Score 29; DB 1; Length 417;  
Best Local Similarity 83.3%; Pred. No. 1.5e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 EKAGGF 6  
:|||||  
Db 183 QKAGGF 188

RESULT 10  
A25567  
phosphoglycerate kinase (EC 2.7.2.3) - mouse  
C:Species: Mus musculus (house mouse)  
C:Date: 30-Jun-1988 #sequence revision 30-Jun-1988 #text\_change 18-Jun-1999  
C:Accession: A25567; A29613; S21393; I48708; I48710  
R:Mori, N.; Singer-Sam, J.; Lee, C.Y.; Riggs, A.D.  
Gene 45, 275-280, 1986  
A:Title: The nucleotide sequence of a cDNA clone containing the entire coding region for  
A:Reference number: A25567; MUID:87106824; PMID:3542714  
A:Accession: A25567  
A:Molecule type: mRNA  
A:Residues: 1-417 <MOR>

A:Cross-references: GB:M15669; NID:g202422; PIDN:AAA70267.1; PID:g202423  
A:Note: the authors translated the codon AAC for residue 56 as Lys  
R:Adra, C.N.; Boer, P.H.; McBurney, M.W.  
Gene 60, 65-74, 1987  
A:Title: Cloning and expression of the mouse pgk-1 gene and the nucleotide sequence of  
A:Reference number: A29613; MUID:88152503; PMID:3440520  
A:Accession: A29613  
A:Molecule type: DNA  
A:Residues: 1-22 <ADR>  
R:Boer, P.H.; Potten, H.; Adra, C.; Jardine, K.; Mllhofer, G.; McBurney, M.  
submitted to the EMBL Data Library, May 1989

A:Description: The mouse pgk-1 promoter induces efficient gene expression: identification  
A:Reference number: S21393  
A:Accession: S21393  
A:Molecule type: DNA  
A:Residues: 1-21 <BOR>  
A:Cross-references: EMBL:X15339; NID:g53671; PIDN:CAA33391.1; PID:g53672  
R:Tamaru, M.; Nagao, Y.; Taira, M.; Tatibana, M.; Masamune, Y.; Nakanishi, Y.  
Biochem. Biophys. Acta 1049, 331-338, 1990  
A:Title: Selective activation of testis-specific genes in cultured rat spermatogenic cells  
A:Reference number: I48708; MUID:90344879; PMID:2166582  
A:Accession: I48708  
A:Status: preliminary; translated from GB/EMBL/DBDJ

A:Molecule type: DNA  
A:Residues: 1-21 <RES>  
A:Cross-references: EMBL:X55309; NID:g57875; PIDN:CAA39013.1; PID:g57876  
R:Boer, P.H.; Potten, H.; Adra, C.N.; Jardine, K.; Mullhofer, G.; McBurney, M.W.  
Biochem. Genet. 28, 299-308, 1990  
A:Title: Polymorphisms in the coding and noncoding regions of murine Pgk-1 alleles.  
A:Reference number: I48710; MUID:90365677; PMID:1975492  
A:Accession: I48710  
A:Status: preliminary; translated from GB/EMBL/DBDJ

A:Molecule type: DNA  
A:Residues: 1-21 <RE2>  
A:Cross-references: EMBL:X15339; NID:g53671; PIDN:CAA33391.1; PID:g53672  
C:Genetics:  
A:Gene: pgk-1  
A:Introns: 22/2

C:Superfamily: phosphoglycerate kinase  
C:Keywords: ATP; gluconeogenesis; glycolysis; phosphotransferase

Query Match 90.6%; Score 29; DB 2; Length 417;  
Best Local Similarity 83.3%; Pred. No. 1.5e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 EKAGGF 6  
:|||||  
Db 183 QKAGGF 188

RESULT 11  
I48074  
phosphoglycerate kinase (EC 2.7.2.3) - Chinese hamster  
C:Species: Cricetulus griseus (Chinese hamster)  
C:Date: 02-Jul-1996 #sequence\_revision 02-Jul-1996 #text\_change 18-Jun-1999  
C:Accession: I48074  
R:Hassan-Walker, A.F.; Morgan, M.J.; Faik, P.  
Somat. Cell Mol. Genet. 21, 75-81, 1995  
A:Title: Characterization of cDNAs coding for glucose phosphate isomerase and phosphoglyc  
efficient variant of CHO-K1.  
A:Reference number: I48073; MUID:95328037; PMID:7604358  
A:Accession: I48074  
A:Status: preliminary; translated from GB/EMBL/DBDJ

A:Molecule type: mRNA  
A:Residues: 1-417 <RES>  
A:Cross-references: EMBL:Z37974; NID:g987047; PIDN:CAA86028.1; PID:g987048  
C:Superfamily: phosphoglycerate kinase  
C:Keywords: ATP; phosphotransferase

Query Match 90.6%; Score 29; DB 2; Length 417;  
Best Local Similarity 83.3%; Pred. No. 1.5e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 EKAGGF 6  
:|||||  
Db 183 QKAGGF 188

RESULT 12  
A33792  
phosphoglycerate kinase (EC 2.7.2.3) - rat  
C:Species: Rattus norvegicus (Norway rat)  
C:Date: 23-Mar-1990 #sequence\_revision 23-Mar-1990 #text\_change 18-Jun-1999  
C:Accession: A33792  
R:Ciccarese, S.; Tommasi, S.; Vonghia, G.  
Biochem. Biophys. Res. Commun. 165, 1337-1344, 1989  
A:Title: Cloning and cDNA sequence of the rat X-chromosome linked phosphoglycerate kinase  
A:Reference number: A33792; MUID:90121246; PMID:2610697  
A:Accession: A33792

A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-417 <CIC>  
A:Cross-references: GB:M31788; NID:g206112; PIDN:AAA41838.1; PID:g206113  
C:Superfamily: phosphoglycerate kinase  
C:Keywords: ATP; phosphotransferase

Query Match 90.6%; Score 29; DB 2; Length 417;  
Best Local Similarity 83.3%; Pred. No. 1.5e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 EKAGGF 6  
:|||||  
Db 183 QKAGGF 188

RESULT 13  
B84702  
hypothetical protein At2g29910 [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 16-Feb-2001  
C:Accession: B84702

R.; Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.; M.; Koo, H.; Mofat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Rallon, L.; euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J. Nature 402, 761-768, 1999

A>Title: Sequence and analysis of chromosome 2 of the plant *Arabidopsis thaliana*.

A;Reference number: A84420; MUID:20083487; PMID:10617197

A;Accession: B84702

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-421 <STO>

A;Cross-references: GB:AE002093; NID:g4567216; PIDN:AAD23631.1; GSPDB:GN00139

C;Genetics:

A;Gene: At2g29910

A;Map position: 2

C;Superfamily: Arabidopsis thaliana hypothetical protein F17J16.30

Query Match 90.6%; Score 29; DB 2; Length 421;  
Best Local Similarity 83.3%; Pred. No. 1.5e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 EKAGGF 6  
Db 265 EKAGGY 270  
|||||

RESULT 14

B87269

hypothetical protein CC0163 [imported] - *Caulobacter crescentus*

C;Species: *Caulobacter crescentus*

C;Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 20-Apr-2001

C;Accession: B87269

R;Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.; B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.B.; Haft, D.H.; Kolon, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M. Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001

A;Title: Complete Genome Sequence of *Caulobacter crescentus*.

A;Reference number: A87249; MUID:21173698; PMID:11259647

A;Accession: B87269

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-457 <STO>

A;Cross-references: GB:AE005673; NID:g13421278; PIDN:AAK22150.1; GSPDB:GN00148

C;Genetics:

A;Gene: CC0163

Query Match 90.6%; Score 29; DB 2; Length 457;  
Best Local Similarity 83.3%; Pred. No. 1.7e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 EKAGGF 6  
Db 86 QKAGGF 91  
|||||

RESULT 15

T28687

hypothetical protein - *Streptomyces coelicolor*

C;Species: *Streptomyces coelicolor*

C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 20-Jun-2000

C;Accession: T28687

R;Parkhill, J.; Bentley, S.D.; Barrell, B.G.; Rajandream, M.A. submitted to the EMBL Data Library, January 1999

A;Reference number: Z20512

A;Accession: T28687

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 1-522 <PAR>

A;Cross-references: EMBL:AL023496; PIDN:CAA18904.1

C;Superfamily: *Escherichia coli* probable transport protein b0511

Query Match 90.6%; Score 29; DB 2; Length 522;  
Best Local Similarity 83.3%; Pred. No. 1.9e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 EKAGGF 6  
Db 229 DXAGGF 234

Search completed: September 18, 2004, 04:30:48  
Job time : 12.2941 secs



Result No.	Score	Query		DB	ID	Description
		Match	Length			
1	32	100.0	97	1	RL28_RICCN	Q99jdl rickettsia
2	32	100.0	97	1	RL28_RICCN	Q99e48 rickettsia
3	32	100.0	1466	1	CA13_HUMAN	P02461 homo sapien
4	29	90.6	300	1	GCST PARTE	P49444 paramycium
5	29	90.6	372	1	PCST STRAW	Q82ji2 streptomyces
6	29	90.6	382	1	OYEA SCHPO	Q09670 schizosacch
7	29	90.6	416	1	PGK1 MOUSE	P09411 mus musculus
8	29	90.6	416	1	PGK2 RAT	P16617 rattus norv
9	29	90.6	416	1	PGK_GRGR	P50310 cricetulus
10	29	90.6	416	1	PGK_HORSE	P00559 equus cabal
11	29	90.6	417	1	PGK1_HUMAN	P00558 homo sapien
12	29	90.6	431	1	KRE2 CANAL	Q00310 candida alb
13	29	90.6	522	1	KTR5 YEAST	P53966 saccharomyc
14	29	90.6	631	1	VG42 HAEIN	P44236 haemophilus
15	29	90.6	636	1	CA13_RAT	P13941 rattus norv
16	29	90.6	1464	1	CA13_MOUSE	P08121 mus musculus
17	28	87.5	93	1	CH10_STRGN	Q8vt59 streptococc
18	28	87.5	293	1	R3P4_HUMAN	Q13868 homo sapien
19	28	87.5	319	1	HSLO CLOPE	Q8xj60 clostridium
20	28	87.5	578	1	VATA_METAC	P28316 cyprinus ca
21	28	87.5	1852	1	CCAS_CYPCA	Q81ij1 methanosarc
22	27	84.4	168	1	DYR_BACSU	P11045 bacillus su
23	27	84.4	301	1	OTC_METTH	O27495 methanobact
24	27	84.4	317	1	CYF_CHLRE	P23577 chlamydomon
25	27	84.4	344	1	DCUP_BRAJA	Q89ek1 bradyrhizob
26	27	84.4	373	1	DMT1_PIG	Q99t01 sus scrofa
27	27	84.4	373	1	Y085_METUA	Q57v50 methanococc
28	27	84.4	379	1	O71A_DROME	Q9rvk5 drosophila
29	27	84.4	379	1	YG36_METWA	Q8pwf3 methanosarc
30	27	84.4	388	1	TRM1_METAC	P48t96 methanosarc
31	27	84.4	393	1	S17P_ARATH	P46283 arabidopsar
32	27	84.4	420	1	CT13_HUMAN	Q9hhp5 homo sapien
33	27	84.4	420	1	CT13_MOUSE	Q8rlg1 mus musculus

28-FEB-2003 (Rel. 41, Last annotation update)  
50S ribosomal protein L28.  
RPMB OR RP099.  
Rickettsia prowazekii.  
Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;  
OC Rickettsiaceae; Rickettsieae; Rickettsia.  
NCBI\_TaxID=782;  
[1]  
SEQUENCE FROM N.A.  
RC STRAIN=Madrid E;  
RX MEDLINE=99039499; PubMed=9823893;  
RA Andersson S.G.E., Zomorodipour A., Andersson J.O., Naelund A.K.,  
SA Sacheritz-Ponten T., Alsmark U.C.M., Podowski R.M., Kurland C.G.;  
RA Eriksson A.-S., Winkler H.H., Kurland C.G.;  
RT "The genome sequence of Rickettsia prowazekii and the origin of  
RT mitochondria.";  
RL Nature 396:133-140 (1998).  
CC -1- SIMILARITY: Belongs to the L28P family of ribosomal proteins.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
CC  
DR EMBL; AJ235270; CAA14569.1; -.  
DR PIR; B71719; B71719.  
DR HAMAP; MF\_00373; -; 1.  
DR InterPro; IPR001383; Ribosomal\_L28.  
DR Pfam; PF00830; Ribosomal\_L28; 1.  
DR TIGRFAMs; TIGR00009; L28; 1.  
DR Ribosomal protein; Complete proteome.  
KW Ribosomal protein; 10953 MW; 96889B2F4E1E28C8 CRC64;  
SQ  
Query Match 100.0%; Score 32; DB 1; Length 97;  
Best Local Similarity 100.0%; Pred. No. 3.9;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 EKAGGF 6  
Dd 60 EKAGGF 65  
|||||  
-----  
RESULT 3  
CAL3 HUMAN  
ID -CAL3 HUMAN STANDARD; PRT; 1466 AA.  
AC P02461; Q15112;  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 01-JAN-1990 (Rel. 13, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Collagen alpha 1(III) chain precursor.  
GN COL3A1.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
[1]  
RN R  
RP SEQUENCE FROM N.A.  
RC TISSUE=Skin fibroblast;  
RX MEDLINE=89350838; PubMed=2764886;  
RA Ala-Kokko L., Kontusaari S., Baldwin C.T., Kuivaniemi H.,  
RA Prockop D.J.;  
RT "Structure of cDNA clones coding for the entire prepro alpha 1 (III)  
RT chain of human type III procollagen. Differences in protein structure  
RT from type I procollagen and conservation of codon preferences.";  
RL Biochem. J. 260:509-516 (1989).  
[2]  
RN R  
RP SEQUENCE OF 149-1225 FROM N.A.  
RX MEDLINE=89386015; PubMed=2780304;  
RA Janeczko R.A., Ramirez F.;  
RT "Nucleotide and amino acid sequences of the entire human alpha 1

RT (III) collagen.";  
RL Nucleic Acids Res. 17:6742-6742 (1989).  
[3]  
RN R  
RP SEQUENCE OF 168-398.  
RX MEDLINE=77134724; PubMed=557335;  
RA Seyer J.M., Kang A.H.;  
RT "Covalent structure of collagen: amino acid sequence of cyanogen  
RT bromide peptides from the amino-terminal segment of type III collagen  
RT of human liver.";  
RL Biochemistry 16:1158-1164 (1977).  
[4]  
RN R  
RP REVISIONS.  
RA Seyer J.M.;  
RL Submitted (DEC-1977) to the PIR data bank.  
[5]  
RN R  
RP SEQUENCE OF 399-727.  
RX MEDLINE=79000343; PubMed=687591;  
RA Seyer J.M., Kang A.H.;  
RT "Covalent structure of collagen: amino acid sequence of five  
RT consecutive CNBr peptides from type III collagen of human liver.";  
RL Biochemistry 17:3404-3411 (1978).  
[6]  
RN R  
RP SEQUENCE OF 728-964.  
RX MEDLINE=80198282; PubMed=6246925;  
RA Seyer J.M., Mainardi C., Kang A.H.;  
RT "Covalent structure of collagen: amino acid sequence of alpha 1  
RT (III)-CBS from type III collagen of human liver.";  
RL Biochemistry 19:1583-1589 (1980).  
[7]  
RN R  
RP SEQUENCE OF 950-1466 FROM N.A.  
RX MEDLINE=88189827; PubMed=3357782;  
RA Mankoo B.S., Dalglish R.;  
RT "Human pro alpha 1(III) collagen: cDNA sequence for the 3' end.";  
RL Nucleic Acids Res. 16:2337-2337 (1988).  
[8]  
RN R  
RP REVISION TO 1184.  
RX MEDLINE=89098346; PubMed=3211760;  
RA Molyneux K., Dalglish R.;  
RT "Human type III collagen, variant, is a cDNA cloning artefact.";  
RL Nucleic Acids Res. 16:11833-11833 (1988).  
[9]  
RN R  
RP SEQUENCE OF 1065-1466 FROM N.A.  
RX MEDLINE=85087944; PubMed=6096827;  
RA Loidl H.R., Brinker J.M., May M., Pihlajaniemi T., Morrow S.,  
RA Rosenbloom J., Myers J.C.;  
RT "Molecular cloning and carboxyl-propeptide analysis of human type III  
RT procollagen.";  
RL Nucleic Acids Res. 12:9383-9394 (1984).  
[10]  
RN R  
RP SEQUENCE OF 965-1200.  
RX MEDLINE=81208139; PubMed=7016180;  
RA Seyer J.M., Kang A.H.;  
RT "Covalent structure of collagen: amino acid sequence of alpha  
RT 1(III)-CB9 from type III collagen of human liver.";  
RL Biochemistry 20:2621-2627 (1981).  
[11]  
RN R  
RP SEQUENCE OF 1176-1466 FROM N.A.  
RX MEDLINE=85157600; PubMed=2579949;  
RA Chu M.-L., Weil D., de Wet W.J., Bernard M.P., Sippola M., Ramirez F.;  
RT "Isolation of cDNA and genomic clones encoding human pro-alpha 1  
RT (III) collagen. Partial characterization of the 3' end region of the  
RT gene.";  
RL J. Biol. Chem. 260:4357-4363 (1985).  
[12]  
RN R  
RP SEQUENCE OF 1161-1200 FROM N.A.  
RX MEDLINE=86187804; PubMed=3754462;  
RA Miskulin M., Dalglish R., Klueve-Beckerman B., Rennard S.I.,  
RA Tolstoshev P., Brantly M., Crystal R.G.;  
RT "Human type III collagen gene expression is coordinately modulated  
RT with the type I collagen genes during fibroblast growth.";  
RL Biochemistry 25:1408-1413 (1986).  
[13]  
RN R  
RP SEQUENCE OF 1-170 FROM N.A.

```

RA Richards A.J., Ward P.N., Narcisi P., Nicholls A.C., Lloyd J.C.,
RA Pope F.M.;
RT "A single base mutation in the gene for type III collagen (COL3A1)
RT converts glycine 847 to glutamic acid in a family with Ehlers-Danlos
RT syndrome type IV. An unaffected family member is mosaic for the
RT mutation.";
RL Hum. Genet. 89:414-418 (1992).
RN [23]
RP VARIANT EDS-IV ASP-1050.
RX MEDLINE=90037070; PubMed=2808425;
RA Tromp G., Kuivaniemi H., Stolle C.A., Pope F.M., Prockop D.J.;
RT "Single base mutation in the type III procollagen gene that converts
RT the codon for glycine 883 to aspartate in a mild variant of
RT Ehlers-Danlos syndrome IV.";
RL J. Biol. Chem. 264:19313-19317 (1989).
RN [24]
RP VARIANT EDS-IV VAL-1077.
RX MEDLINE=91374480; PubMed=1895316;
RA Richards A.J., Lloyd J.C., Ward P.N., de Paep A., Narcisi P.,
RA Pope F.M.;
RT "Characterisation of a glycine to valine substitution at amino acid
RT position 910 of the triple helical region of type III collagen in a
RT patient with Ehlers-Danlos syndrome type IV.";
RL J. Med. Genet. 28:458-463 (1991).
RN [25]
RP VARIANT EDS-IV GLU-1173.
RX MEDLINE=93022543; PubMed=1357232;
RA Johnson P.H., Richards A.J., Pope F.M., Hopkinson D.A.;
Query Match 100.0%; Score 32; DB 1; Length 1466;
Seq Local Similarity 100.0%; Pred. No. 51;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps
QY 1 EKAGGF 6
DB 1211 EKAGGF 1216
|||||
|||||

RESULT 4
ID_P2C PARTE STANDARD; PRT; 300 AA.
AC P49444;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Protein phosphatase 2C (EC 3.1.3.16) (PP2C).
OS Paramécium tetraurelia.
OC Eukaryota; Alveolata; Ciliophora; Oligohymenophorea; Peniculida;
OC Paramécium.
ON NCBI_TaxID=5888;
OX [1]
XP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RP STRAIN=Stock 51;
RX MEDLINE=95105156; PubMed=7806499;
RA Klump S., Hanke C., Donella-Deana A., Beyer A., Kellner R.,
RA Pinna L.A., Schultz J.E.;
RT "A membrane-bound protein phosphatase type 2C from Paramécium
RT tetraurelia. Purification, characterization, and cloning.";
RL J. Biol. Chem. 269:32774-32780 (1994).
RN [26]
CC -! FUNCTION: Enzyme with a broad specificity.
CC -! CATALYTIC ACTIVITY: A phosphoprotein + H(2)O = a protein +
CC phosphate.
CC -! COFACTOR: Binds 2 magnesium or manganese ions (By similarity).
CC -! SUBCELLULAR LOCATION: Membrane-bound.
CC -! PTM: The N-terminus is blocked.
CC -! SIMILARITY: Belongs to the PP2C family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collabor
CC between the Swiss Institute of Bioinformatics and the EMBL outstatu
CC the European Bioinformatics Institute. There are no restrictions on
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commu
CC entities requires a license agreement (See http://www.isb-sib.ch/anno

```

CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).

DR EMBL; Z36985; CAA85448.1; -;  
DR PIR; A55804; A55804.  
DR HSSP; P35813; 1A6Q.  
DR InterPro; IPR001932; PP2C-like.  
DR InterPro; IPR002222; PP2C.  
DR Pfam; PF00481; PP2C; 1.  
DR SMART; SM00331; PP2C SIG; 1.  
DR SMART; SM00332; PP2CC; 1.  
DR PROSITE; PS01032; PP2C; 1.  
KW Hydroxylase; Magnesium; Membrane.  
FT METAL 37 38 MANGANESE 1 (BY SIMILARITY).  
FT METAL 38 38 MANGANESE 1 (BY SIMILARITY).  
FT METAL 57 57 MANGANESE 1 AND 2 (BY SIMILARITY).  
FT METAL 237 237 MANGANESE 2 (BY SIMILARITY).  
FT METAL 289 289 MANGANESE 2 (BY SIMILARITY).  
SQ SEQUENCE 300 AA; 33739 MW; EC0318BAFF7724EC CRC64;

Query Match 90.6%; Score 29; DB 1; Length 300;  
Best Local Similarity 83.3%; Pred. No. 52;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 EKAGGF 6  
Db 175 ERAGGF 180

RESULT 5  
ID GST\_STRAW STANDARD; PRT; 372 AA.  
AC Q82J12;  
DT 10-OCT-2003 (Rel. 42, Created)  
DT 10-OCT-2003 (Rel. 42, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Aminomethyltransferase (EC 2.1.2.10) (Glycine cleavage system T  
protein).  
DE GCVT OR SAV2773.  
OS Streptomyces avermitilis.  
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
OC Streptomycineae; Streptomycetaceae; Streptomyces.  
OX NCBI\_TaxID=33903;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;  
MEDLINE=21477403; PubMed=11572948;  
RA Omura S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,  
RA Shinose M., Takahashi Y., Horikawa H., Nakazawa H., Osone T.,  
RA Kikuchi H., Shiba T., Sakaki Y., Hattori M.,  
RA "Genome sequence of an industrial microorganism Streptomyces  
RT avermitilis: deducing the ability of producing secondary  
RT metabolites";  
RL Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220(2001).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;  
MEDLINE=22608306; PubMed=12692562;  
RA Ikeda H., Ishikawa J., Hanamoto A., Shinose M., Kikuchi H., Shiba T.,  
RA Sakaki Y., Hattori M., Omura S.,  
RA "Complete genome sequence and comparative analysis of the industrial  
RT microorganism Streptomyces avermitilis";  
RL Nat. Biotechnol. 21:528-531(2003).  
CC -!- FUNCTION: The glycine cleavage system catalyzes the degradation of  
glycine (By similarity).  
CC -!- CATALYTIC ACTIVITY: (6S)-tetrahydrofolate + S-  
aminomethyldihydropyrimidinylprotein = (6R)-5,10-  
methylenetetrahydrofolate + NH(3) + dihydropyrimidinylprotein.  
CC -!- SUBUNIT: The glycine cleavage system is composed of four proteins:  
P, T, L and H (By similarity).  
CC -!- SIMILARITY: Belongs to the GCVT family.

Query Match 90.6%; Score 29; DB 1; Length 372;  
Best Local Similarity 83.3%; Pred. No. 63;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 EKAGGF 6  
Db 130 ERAGGF 135

RESULT 6  
ID OYEA\_SCHPO STANDARD; PRT; 382 AA.  
AC Q09670;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Putative NADPH dehydrogenase C5H10.04 (EC 1.6.99.1) (old yellow enzyme  
homolog).  
DE SPACSH10.04.  
OS Schizosaccharomyces pombe (Fission yeast).  
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
OC Schizosaccharomycetes.  
OX NCBI\_TaxID=4896;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX STRAIN=972;  
MEDLINE=21848401; PubMed=11859360;  
RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,  
RA Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,  
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,  
RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,  
RA Gentles S., Gobie A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,  
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagsels K.,  
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,  
RA Money P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,  
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,  
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,  
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,  
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,  
RA Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,  
RA Weijtens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,  
RA Gabel C., Fuchs M., Fritz C., Holzer E., Reinhardt R., Pohl T.M.,  
RA Borzym K., Langer I., Beck A., Leirach H., Wambutt R., Furnelle B.,  
RA Eger P., Zimmermann W., Wedler H., Wambutt R., Mottier S.,  
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,  
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,  
RA Lucas M., Rochet M., Gallard C., Tallada V.A., Garzon A., Thode G.,  
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,  
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,  
RA Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,  
RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.,  
RT "The genome sequence of Schizosaccharomyces pombe";  
RL Nature 415:871-880(2002).  
CC -!- CATALYTIC ACTIVITY: NADPH + acceptor = NADP(+) + reduced acceptor.  
CC -!- COFACTOR: FMN.  
CC -!- SUBUNIT: Homodimer or heterodimer (By similarity).

QY 1 EKAGGF 6  
Db 130 ERAGGF 135

RESULT 6  
ID OYEA\_SCHPO STANDARD; PRT; 382 AA.  
AC Q09670;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Putative NADPH dehydrogenase C5H10.04 (EC 1.6.99.1) (old yellow enzyme  
homolog).  
DE SPACSH10.04.  
OS Schizosaccharomyces pombe (Fission yeast).  
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
OC Schizosaccharomycetes.  
OX NCBI\_TaxID=4896;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX STRAIN=972;  
MEDLINE=21848401; PubMed=11859360;  
RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,  
RA Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,  
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,  
RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,  
RA Gentles S., Gobie A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,  
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagsels K.,  
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,  
RA Money P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,  
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,  
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,  
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,  
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,  
RA Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,  
RA Weijtens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,  
RA Gabel C., Fuchs M., Fritz C., Holzer E., Reinhardt R., Pohl T.M.,  
RA Borzym K., Langer I., Beck A., Leirach H., Wambutt R., Furnelle B.,  
RA Eger P., Zimmermann W., Wedler H., Wambutt R., Mottier S.,  
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,  
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,  
RA Lucas M., Rochet M., Gallard C., Tallada V.A., Garzon A., Thode G.,  
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,  
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,  
RA Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,  
RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.,  
RT "The genome sequence of Schizosaccharomyces pombe";  
RL Nature 415:871-880(2002).  
CC -!- CATALYTIC ACTIVITY: NADPH + acceptor = NADP(+) + reduced acceptor.  
CC -!- COFACTOR: FMN.  
CC -!- SUBUNIT: Homodimer or heterodimer (By similarity).

QY 1 EKAGGF 6  
Db 130 ERAGGF 135

RESULT 6  
ID OYEA\_SCHPO STANDARD; PRT; 382 AA.  
AC Q09670;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Putative NADPH dehydrogenase C5H10.04 (EC 1.6.99.1) (old yellow enzyme  
homolog).  
DE SPACSH10.04.  
OS Schizosaccharomyces pombe (Fission yeast).  
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
OC Schizosaccharomycetes.  
OX NCBI\_TaxID=4896;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX STRAIN=972;  
MEDLINE=21848401; PubMed=11859360;  
RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,  
RA Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,  
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,  
RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,  
RA Gentles S., Gobie A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,  
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagsels K.,  
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,  
RA Money P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,  
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,  
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,  
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,  
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,  
RA Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,  
RA Weijtens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,  
RA Gabel C., Fuchs M., Fritz C., Holzer E., Reinhardt R., Pohl T.M.,  
RA Borzym K., Langer I., Beck A., Leirach H., Wambutt R., Furnelle B.,  
RA Eger P., Zimmermann W., Wedler H., Wambutt R., Mottier S.,  
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,  
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,  
RA Lucas M., Rochet M., Gallard C., Tallada V.A., Garzon A., Thode G.,  
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,  
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,  
RA Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,  
RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.,  
RT "The genome sequence of Schizosaccharomyces pombe";  
RL Nature 415:871-880(2002).  
CC -!- CATALYTIC ACTIVITY: NADPH + acceptor = NADP(+) + reduced acceptor.  
CC -!- COFACTOR: FMN.  
CC -!- SUBUNIT: Homodimer or heterodimer (By similarity).

QY 1 EKAGGF 6  
Db 130 ERAGGF 135

RESULT 6  
ID OYEA\_SCHPO STANDARD; PRT; 382 AA.  
AC Q09670;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Putative NADPH dehydrogenase C5H10.04 (EC 1.6.99.1) (old yellow enzyme  
homolog).  
DE SPACSH10.04.  
OS Schizosaccharomyces pombe (Fission yeast).  
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
OC Schizosaccharomycetes.  
OX NCBI\_TaxID=4896;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX STRAIN=972;  
MEDLINE=21848401; PubMed=11859360;  
RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,  
RA Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,  
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,  
RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,  
RA Gentles S., Gobie A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,  
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagsels K.,  
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,  
RA Money P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,  
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,  
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,  
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,  
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,  
RA Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,  
RA Weijtens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,  
RA Gabel C., Fuchs M., Fritz C., Holzer E., Reinhardt R., Pohl T.M.,  
RA Borzym K., Langer I., Beck A., Leirach H., Wambutt R., Furnelle B.,  
RA Eger P., Zimmermann W., Wedler H., Wambutt R., Mottier S.,  
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,  
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,  
RA Lucas M., Rochet M., Gallard C., Tallada V.A., Garzon A., Thode G.,  
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,  
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,  
RA Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,  
RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.,  
RT "The genome sequence of Schizosaccharomyces pombe";  
RL Nature 415:871-880(2002).  
CC -!- CATALYTIC ACTIVITY: NADPH + acceptor = NADP(+) + reduced acceptor.  
CC -!- COFACTOR: FMN.  
CC -!- SUBUNIT: Homodimer or heterodimer (By similarity).

QY 1 EKAGGF 6  
Db 130 ERAGGF 135

RESULT 6  
ID OYEA\_SCHPO STANDARD; PRT; 382 AA.  
AC Q09670;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Putative NADPH dehydrogenase C5H10.04 (EC 1.6.99.1) (old yellow enzyme  
homolog).  
DE SPACSH10.04.  
OS Schizosaccharomyces pombe (Fission yeast).  
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
OC Schizosaccharomycetes.  
OX NCBI\_TaxID=4896;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX STRAIN=972;  
MEDLINE=21848401; PubMed=11859360;  
RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,  
RA Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,  
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,  
RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,  
RA Gentles S., Gobie A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,  
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagsels K.,  
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,  
RA Money P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,  
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,  
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,  
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,  
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,  
RA Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,  
RA Weijtens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,  
RA Gabel C., Fuchs M., Fritz C., Holzer E., Reinhardt R., Pohl T.M.,  
RA Borzym K., Langer I., Beck A., Leirach H., Wambutt R., Furnelle B.,  
RA Eger P., Zimmermann W., Wedler H., Wambutt R., Mottier S.,  
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,  
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,  
RA Lucas M., Rochet M., Gallard C., Tallada V.A., Garzon A., Thode G.,  
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,  
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,  
RA Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,  
RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.,  
RT "The genome sequence of Schizosaccharomyces pombe";  
RL Nature 415:871-880(2002).  
CC -!- CATALYTIC ACTIVITY: NADPH + acceptor = NADP(+) + reduced acceptor.  
CC -!- COFACTOR: FMN.  
CC -!- SUBUNIT: Homodimer or heterodimer (By similarity).

QY 1 EKAGGF 6  
Db 130 ERAGGF 135

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; Z49811; CAA89954.1; --  
 PIR; S55482; S55482.  
 HSP; Q02899; LOYC.  
 GeneDB\_Spombe; SPAC5H10.04; --  
 InterPro; IPR001155; Oxidored FMN.  
 Pfam; PF00724; oxidored FMN; I.  
 KW Hypothetical protein; Oxidoreductase; NADP; Flavoprotein; FMN;  
 Multigene family.  
 SQ SEQUENCE 382 AA; 43813 MW; 3F7939599CA167D1 CRC64;

Query Match 90.6%; Score 29; DB 1; Length 382;  
 Best Local Similarity 83.3%; Pred. No. 65;  
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 EXAGGF 6  
 DB 66 EKSGGF 71

## RESULT 7

PGK1\_MOUSE  
 ID PGK1\_MOUSE STANDARD; PRT; 416 AA.  
 AC P09411;  
 DT 01-MAR-1989 (Rel. 10, Created)  
 DT 01-OCT-1989 (Rel. 12, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Phosphoglycerate kinase 1 (EC 2.7.2.3).  
 GN PGK1 OR PGK-1.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=87106824; PubMed=3542714;  
 RA Mori N., Singer-Sam J., Lee C.-Y., Riggs A.D.;  
 RT "The nucleotide sequence of a cDNA clone containing the entire coding  
 region for mouse X-chromosome-linked phosphoglycerate kinase.";  
 RL Gene 45:275-280(1986).  
 RN [2]  
 RP DISCUSSION OF SEQUENCE.  
 RX MEDLINE=86275275; PubMed=3525226;  
 RA Mori N., Singer-Sam J., Riggs A.D.;  
 RT "Evolutionary conservation of the substrate-binding cleft of  
 phosphoglycerate kinases.";  
 RL FEBS Lett. 204:313-317(1986).  
 RN [3]  
 RP SEQUENCE OF 1-20 FROM N.A.  
 RX MEDLINE=88152503; PubMed=3440520;  
 RA Adra C.N., Boer P.H., McBurney M.W.;  
 RT "Cloning and expression of the mouse pgk-1 gene and the nucleotide  
 sequence of its promoter.";  
 RL Gene 60:65-74(1987).  
 RN [4]  
 RP SEQUENCE OF 1-20 FROM N.A.  
 RX MEDLINE=90344879; PubMed=2166582;  
 RA Tanaru M., Ngao Y., Taira M., Tatibana M., Masamune Y., Nakanishi Y.;  
 RT "Selective activation of testis-specific genes in cultured rat  
 spermatogenic cells.";  
 RL Biochim. Biophys. Acta 1049:331-338(1990).  
 RN [5]  
 RP SEQUENCE OF 1-20 FROM N.A.  
 RX MEDLINE=90365677; PubMed=1975492;  
 RA Boer P.H., Potten H., Adra C.N., Jardine K., Mullhofer G.,  
 RA McBurney M.W.;

RT "Polymorphisms in the coding and noncoding regions of murine Pgk-1  
 alleles.";  
 RL Biochem. Genet. 28:299-299(1990).  
 CC -!- CATALYTIC ACTIVITY: ATP + 3-phospho-D-glycerate = ADP + 3-  
 phospho-D-glyceroyl phosphate.  
 CC -!- PATHWAY: Second phase of glycolysis; second step.  
 CC -!- SUBUNIT: Monomer.  
 CC -!- SIMILARITY: Belongs to the phosphoglycerate kinase family.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; M15668; AAA70267.1; --  
 EMBL; M18735; AAA39919.1; --  
 EMBL; X55309; CAA39013.1; --  
 EMBL; X15339; CAA33391.1; --  
 PIR; A25567; A25567.  
 DR HSP; P00560; 10PG.  
 DR SWISS-2DPAGE; P09411; MOUSE.  
 DR MGD; MGI:97555; Pgk1.  
 DR InterPro; IPR001576; PGK.  
 DR Pfam; PF00162; PGK; 1.  
 DR PRINTS; PR00477; PHGLYCKINASE.  
 DR PROSITE; PS00111; GLYCERATE KINASE; 1.  
 KW Transferase; Kinase; Multigene family; Glycolysis.  
 FT INIT MET 0  
 FT SEQUENCE 416 AA; 44405 MW; 8B879E5AD7E69E15 CRC64;

Query Match 90.6%; Score 29; DB 1; Length 416;  
 Best Local Similarity 83.3%; Pred. No. 70;  
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 EXAGGF 6  
 DB 182 QKAGGF 187

## RESULT 8

PGK2\_RAT  
 ID PGK2\_RAT STANDARD; PRT; 416 AA.  
 AC P16617;  
 DT 01-AUG-1990 (Rel. 15, Created)  
 DT 01-AUG-1990 (Rel. 15, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Phosphoglycerate kinase, testis specific (EC 2.7.2.3).  
 GN PGK2 OR PGK-2.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=90121246; PubMed=2610697;  
 RA Ciccarese S., Tommasi S., Vonghia G.;  
 RT "Cloning and cDNA sequence of the rat X-chromosome linked  
 phosphoglycerate kinase.";  
 RL Biochem. Biophys. Res. Commun. 165:1337-1344(1989).  
 CC -!- CATALYTIC ACTIVITY: ATP + 3-phospho-D-glycerate = ADP + 3-  
 phospho-D-glyceroyl phosphate.  
 CC -!- PATHWAY: Second phase of glycolysis; second step.  
 CC -!- SUBUNIT: Monomer.  
 CC -!- SIMILARITY: Belongs to the phosphoglycerate kinase family.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
CC EMBL; M31789; AAA41838.1; -;  
CC PIR; A33792; A33792.  
CC HSSP; P00560; IQPG.  
DR InterPro; IPR001576; PGK.  
DR Pfam; PF00162; PGK; 1.  
DR PRINTS; PR00477; PHGLYCKINASE.  
DR PROSITE; PS00111; GLYCERATE KINASE; 1.  
KW Transferase; Kinase; Multigene family; Glycolysis.  
FT INIT MET 0  
SQ SEQUENCE 416 AA; 44423 MW; 991BEC54432FA8DB CRC64;  
  
Query Match 90.6%; Score 29; DB 1; Length 416;  
Best Local Similarity 83.3%; Pred. No. 70;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 EKAGGF 6  
Db 182 QKAGGF 187

RESULT 9  
PGK\_CRIGR STANDARD; PRT; 416 AA.  
AC P50310;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Phosphoglycerate kinase (EC 2.7.2.3).  
OS Cricetus.  
GN Cricetus griseus (Chinese hamster).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Cricetinae;  
OC Cricetus.  
OX NCBI\_TaxID=10029;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Ovary;  
RX MEDLINE=95328037; PubMed=7604358;  
RA Hassan A.F., Morgan M.J., Paik P.;  
RT "Characterization of cDNAs coding for glucose phosphate isomerase and phosphoglycerate kinase in Chinese hamster ovary cell line CHO-K1 and identification of defects in Ki.1.7, a glycolysis-deficient variant of CHO-K1.";  
RL Sonat. Cell Mol. Genet. 21:75-81(1995).  
CC -1- CATALYTIC ACTIVITY: ATP + 3-phospho-D-glycerate = ADP + 3-phospho-D-glyceroyl phosphate.  
CC -1- PATHWAY: Second phase of glycolysis; second step.  
CC -1- SUBUNIT: Monomer (By similarity).  
CC -1- SIMILARITY: Belongs to the phosphoglycerate kinase family.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
CC EMBL; Z37974; CAA86028.1; -;  
CC PIR; I48074; I48074.  
CC HSSP; P00560; IQPG.  
DR InterPro; IPR001576; PGK.  
DR Pfam; PF00162; PGK; 1.  
DR PRINTS; PR00477; PHGLYCKINASE.  
DR PROSITE; PS00111; GLYCERATE KINASE; 1.  
KW Transferase; Kinase; Glycolysis.  
FT INIT MET 0  
SQ SEQUENCE 416 AA; 44431 MW; 7DCD55E0940242C5 CRC64;  
  
Query Match 90.6%; Score 29; DB 1; Length 416;  
Best Local Similarity 83.3%; Pred. No. 70;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Best Local Similarity 83.3%; Pred. No. 70;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 EKAGGF 6  
Db 182 QKAGGF 187

RESULT 10  
PGK\_HORSE STANDARD; PRT; 416 AA.  
AC P00559;  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 21-JUL-1986 (Rel. 01, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Phosphoglycerate kinase (EC 2.7.2.3).  
OS Equus caballus (Horse).  
GN Equus caballus (Horse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Euthera; Perissodactyla; Equidae; Equus.  
OX NCBI\_TaxID=9796;  
RN [1]  
RP SEQUENCE.  
RX MEDLINE=82030789; PubMed=7287713;  
RA Merrett M.;  
RT "Primary structure of 3-phosphoglycerate kinase from horse muscle. II. Amino acid sequence of cyanogen bromide peptides CBI-CB4 and CBI-CB14, sequence of methionine-containing regions, and complete sequence of the enzyme.";  
RL J. Biol. Chem. 256:10293-10305(1981).  
RN [2]  
RP X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS).  
RC TISSUE=Muscle;  
RX MEDLINE=79199779; PubMed=450128;  
RA Banks R.D., Blake C.C.F., Evans P.R., Haser R., Rice D.W., Hardy G.W., Merrett M., Phillips A.W.;  
RT "Sequence, structure and activity of phosphoglycerate kinase: a possible hinge-bending enzyme.";  
RL Nature 279:773-777(1979).  
CC -1- CATALYTIC ACTIVITY: ATP + 3-phospho-D-glycerate = ADP + 3-phospho-D-glyceroyl phosphate.  
CC -1- PATHWAY: Second phase of glycolysis; second step.  
CC -1- SUBUNIT: Monomer.  
CC -1- SIMILARITY: Belongs to the phosphoglycerate kinase family.  
CC PDB; 2PGK; 17-FEB-84.  
DR InterPro; IPR001576; PGK.  
DR Pfam; PF00162; PGK; 1.  
DR PRINTS; PR00477; PHGLYCKINASE.  
DR PROSITE; PS00111; GLYCERATE KINASE; 1.  
KW Transferase; Kinase; Acetylation; Glycolysis; 3D-structure.  
FT MOD\_RES 1 185  
FT DOMAIN 1 185  
FT DOMAIN 186 189  
FT DOMAIN 190 416  
FT DOMAIN 416 416  
FT BINDING 219 219  
FT BINDING 212 214  
FT BINDING 236 238  
FT BINDING 339 341  
FT BINDING 343 343  
SQ SEQUENCE 416 AA; 44471 MW; A4B926A6FE0B0FDA CRC64;  
  
Query Match 90.6%; Score 29; DB 1; Length 416;  
Best Local Similarity 83.3%; Pred. No. 70;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

RESULT 11  
PGK1\_HUMAN

ID PGK1 HUMAN STANDARD; PRT; 417 AA.  
AC P00556;  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 21-JUL-1986 (Rel. 01, Last sequence update)  
DE 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Phosphoglycerate kinase 1 (EC 2.7.2.3) (Primer recognition protein 2)  
GN (PRP 2).  
GN PGK1 OR PGKA.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Liver;  
RX MEDLINE=83169680; PubMed=6188151;  
RA Michelson A.M., Markham A.F., Orkin S.H.;  
RT "Isolation and DNA sequence of a full-length cDNA clone for human X  
RT chromosome-encoded phosphoglycerate kinase.";  
RL Proc. Natl. Acad. Sci. U.S.A. 80:472-476(1983).  
RN [2]  
RP SEQUENCE OF 1-416 FROM N.A.  
RX MEDLINE=86016816; PubMed=2995995;  
RA Michelson A.M., Blake C.C., Evans S.T., Orkin S.H.;  
RT "Structure of the human phosphoglycerate kinase gene and the intron-  
RT mediated evolution and dispersal of the nucleotide-binding domain.";  
RL Proc. Natl. Acad. Sci. U.S.A. 82:6965-6969(1985).  
RN [3]  
RP SEQUENCE.  
RC TISSUE=Erythrocyte;  
RX MEDLINE=80227775; PubMed=7391027;  
RA Huang I.-Y., Welch C.D., Yoshida A.;  
RT "Complete amino acid sequence of human phosphoglycerate kinase.  
RT Cyanogen bromide peptides and complete amino acid sequence.";  
RL J. Biol. Chem. 255:6412-6420(1980).  
RN [4]  
RP PARTIAL SEQUENCE.  
RC TISSUE=Placenta;  
RX MEDLINE=90216667; PubMed=2324090;  
RA Jindal H.K., Vishwanatha J.K.;  
RT "Functional identity of a primer recognition protein as  
RT phosphoglycerate kinase.";  
RL J. Biol. Chem. 265:6540-6543(1990).  
RN [5]  
RP SEQUENCE OF 1-20 FROM N.A.  
RX MEDLINE=85155507; PubMed=6099325;  
RA Singer-Sam J., Keith D.H., Tanl K., Simmer R.L., Shively L.,  
RA Lindsay S., Yoshida A., Riggs A.D.;  
RT "Sequence of the promoter region of the gene for human X-linked 3-  
RT phosphoglycerate kinase.";  
RL Gene 32:409-417(1984).  
RN [6]  
RP SEQUENCE OF 1-13 FROM N.A.  
RX MEDLINE=90049205; PubMed=2814502;  
RA Pfeifer G.P., Steigewald S.D., Mueller P.R., Wold B., Riggs A.D.;  
RT "Genomic sequencing and methylation analysis by ligation mediated  
RT PCR.";  
RL Science 246:810-813(1989).  
RN [7]  
RP REVIEW ON VARIANTS.  
RX MEDLINE=97230014; PubMed=9075577;  
RA Yoshida A.;  
RT "Hematologically important mutations: molecular abnormalities of  
RT phosphoglycerate kinase.";  
RL Blood Cells Mol. Dis. 22:265-267(1996).  
RN [8]  
RP VARIANT CHRONIC HEMOLYTIC ANEMIA LYS-190 DEL.  
RX MEDLINE=96230923; PubMed=8673469;  
RA Yoshida A., Twele T.W., Dave V., Beutler E.;  
RT "Molecular abnormality of a phosphoglycerate kinase variant  
RT (PGK-Alabama).";  
RL Blood Cells Mol. Dis. 21:179-181(1995).  
RN [9]  
RP VARIANT CHRONIC HEMOLYTIC ANEMIA/MENTAL RETARDATION VAL-163, AND  
RP VARIANT RHADOMYOLYSIS ASN-314.  
RX MEDLINE=94318968; PubMed=8043870;  
RA Cohen-Solal M., Valentin C., Plassa F., Guillemin G., Danze F.,  
RA Jaisson P., Rosa R.;  
RT "Identification of new mutations in two phosphoglycerate kinase (PGK)  
RT variants expressing different clinical syndromes: PGK Creteil and PGK  
RT Amiens.";  
RL Blood 84:898-903(1994).  
RN [10]  
RP VARIANT CHRONIC HEMOLYTIC ANEMIA ALA-251.  
RX MEDLINE=96201344; PubMed=8615693;  
RA Ookawara T., Dave V., Willems P., Martin J.J., de Barsey T.,  
RA Mathys E., Yoshida A.;  
RT "Retarded and aberrant splicing caused by single exon mutation in a  
RT phosphoglycerate kinase variant.";  
RL Arch. Biochem. Biophys. 327:35-40(1996).  
RN [11]  
RP VARIANT CHRONIC HEMOLYTIC ANEMIA VAL-284.  
RX MEDLINE=98415729; PubMed=9744480;  
RA Valentin C., Birgens H., Craescu C.T., Broedum-Nielsen K.,  
RA Cohen-Solal M.;  
RT "A phosphoglycerate kinase mutant (PGK Herlev; D285V) in a Danish  
RT patient with isolated chronic hemolytic anemia: mechanism of mutation  
RT and structure-function relationships.";  
RL Hum. Mutat. 12:280-287(1998).  
RN [12]  
RP VARIANT CONGENITAL NONSPHEROCYTIC ANEMIA PRO-87.  
RX MEDLINE=91159642; PubMed=2001457;  
RA Maeda M., Yoshida A.;  
RT "Molecular defect of a phosphoglycerate kinase variant (PGK-Matsue)  
RT associated with hemolytic anemia: Leu-->Pro substitution caused by  
RT T/A-->C/G transition in exon 3.";  
RL Blood 77:1348-1352(1991).  
RN [13]  
RP VARIANT CHRONIC HEMOLYTIC ANEMIA ARG-315.  
RX MEDLINE=92265933; PubMed=1586722;  
RA Maeda M., Bawle E.V., Kulkarni R., Beutler E., Yoshida A.;  
RT "Molecular abnormalities of a phosphoglycerate kinase variant  
RT generated by spontaneous mutation.";  
RL Blood 79:2759-2762(1992).  
RN [14]  
RP VARIANT MUNCHEN ASN-267.  
RX MEDLINE=80227776; PubMed=7391028;  
RA Fujii H., Krietsch W.K.G., Yoshida A.;  
RT "A single amino acid substitution (Asp leads to Asn) in a  
RT phosphoglycerate kinase variant (PGK Munchen) associated with enzyme  
RT deficiency.";  
RL J. Biol. Chem. 255:6421-6423(1980).  
RN [15]  
RP VARIANT MUNCHEN ASN-267, AND VARIANT ASN-351.  
RX MEDLINE=81069227; PubMed=7440217;  
RA Huang I.-Y., Fujii H., Yoshida A.;  
RT "Structure and function of normal and variant human phosphoglycerate  
RT kinase.";  
RL Hemoglobin 4:601-609(1980).  
RN [16]  
RP VARIANT CHRONIC HEMOLYTIC ANEMIA VAL-157.  
RX MEDLINE=92190498; PubMed=1547346;  
RA Fujii H., Kanno H., Hirono A., Shionura T., Miwa S.;  
RT "A single amino acid substitution (157 Gly-->Val) in a  
RT phosphoglycerate kinase variant (PGK Shizuoka) associated with  
RT chronic hemolysis and myoglobinuria.";  
RL Blood 79:1582-1585(1992).  
RN [17]  
RP VARIANT CHRONIC NONSPHEROCYTIC HEMOLYTIC ANEMIA MET-265.  
RX MEDLINE=81223926; PubMed=6941312;  
RA Fujii H., Chen S.-H., Akatsuka J., Miwa S., Yoshida A.;  
RT "Use of cultured lymphoblastoid cells for the study of abnormal  
RT enzymes: molecular abnormality of a phosphoglycerate kinase variant  
RT associated with hemolytic anemia.";  
RL Proc. Natl. Acad. Sci. U.S.A. 78:2587-2590(1981).  
RN [18]



RP VARIANT CHRONIC HEMOLYTIC ANEMIA PRO-205.  
 RX MEDLINE=81054987; PubMed=6933565;  
 RA Fujii H., Yoshida A.;  
 RT "Molecular abnormality of phosphoglycerate kinase-Uppsala associated  
 with chronic nonspherocytic hemolytic anemia.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 77:5461-5465(1980).  
 CC -!- FUNCTION: In addition to its role as a glycolytic enzyme, it  
 seems that PGK-1 acts as a polymerase alpha cofactor protein  
 (primer recognition protein).  
 CC -!- CATALYTIC ACTIVITY: ATP + 3-phospho-D-glycerate = ADP + 3-  
 phospho-D-glycerol phosphate.  
 CC -!- PATHWAY: Second phase of glycolysis; second step.  
 CC -!- SUBUNIT: Monomer.  
 CC -!- DISEASE: Defects in PGK1 are generally associated with chronic  
 hemolytic anemia [MIM:311800]; although it can be accompanied by  
 either mental retardation or muscular disease (rhabdomyolysis).  
 CC -!- SIMILARITY: Belongs to the phosphoglycerate kinase family.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation  
 at the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; V00572; CAA23835.1; -.  
 CC EMBL; M11968; AAA60079.1; -.  
 CC EMBL; M11958; AAA60079.1; JOINED.  
 CC EMBL; M11959; AAA60079.1; JOINED.  
 CC EMBL; M11960; AAA60079.1; JOINED.  
 CC EMBL; M11961; AAA60079.1; JOINED.  
 CC EMBL; M11962; AAA60079.1; JOINED.  
 CC EMBL; M11963; AAA60079.1; JOINED.  
 CC EMBL; M11964; AAA60079.1; JOINED.  
 CC EMBL; M11965; AAA60079.1; JOINED.  
 CC EMBL; M11966; AAA60079.1; JOINED.  
 CC EMBL; M11967; AAA60079.1; JOINED.  
 CC EMBL; L00160; AAA60078.1; -.  
 CC EMBL; M34017; AAA60103.1; -.  
 CC HSP; P00560; 1QPG.  
 CC Aarhus/Ghent-2DPG; 3308; NEPHGE.  
 CC Genew; HGNC:8896; PGK1.  
 CC GK; P00558; -.  
 CC MTM; 311800; -.  
 CC GO; GO:0004618; F:phosphoglycerate kinase activity; TAS.  
 CC InterPro; IPR001576; PGK.  
 CC Pfam; PF00162; PGK; 1.  
 CC PRINTS; PR00477; PGKLYCKINASE.  
 CC PROSITE; PS00111; POLYGLYCERATE\_KINASE; 1.  
 CC Transferrase; Kinase; Multigene family; Glycolysis; Acetylation;  
 KW Disease mutation; Polymorphism; Hereditary hemolytic anemia.  
 FT INIT MET 0  
 FT MOD\_RES 1 1 ACETYLATION.  
 FT  
 Query Match 90.6%; Score 29; DB 1; Length 417;  
 Best Local Similarity 83.3%; Pred. No. 71;  
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 EKAGGF 6  
 DB 182 QKAGGF 187  
 :|||||  
 RESULT 12  
 KRE2 CANAL STANDARD; PRT; 431 AA.  
 ID KRE2 CANAL  
 AC Q00310.  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Glycolipid 2-alpha-mannosyltransferase (EC 2.4.1.131) (Alpha-1,2-  
 mannosyltransferase).  
 DE

GN KRE2 OR MNT1.  
 OS Candida albicans (Yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; mitosporic Saccharomycetales; Candida.  
 OX NCBI\_TaxID=5476;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CAI-4;  
 RX MEDLINE=98301633; PubMed=9636208;  
 RA Burman E.I.; Westwater C.; Hube B.; Brown A.J.; Odds F.C.;  
 RA Gow N.A.R.;  
 RT "Molecular analysis of Camtlp, a mannosyl transferase important for  
 adhesion and virulence of Candida albicans.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 95:7670-7675(1998).  
 CC -!- FUNCTION: REQUIRED FOR THE ATTACHMENT OF THE THIRD MANNOSE  
 RESIDUE OF O-LINKED SACCHARIDES (BY SIMILARITY).  
 CC -!- CATALYTIC ACTIVITY: Transfers an alpha-D-mannosyl residue from  
 GDP-mannose into lipid-linked oligosaccharide, forming an alpha-  
 1,2-D-mannosyl-D-mannose linkage.  
 CC -!- PATHWAY: Glycosylation.  
 CC -!- SUBCELLULAR LOCATION: Type II membrane protein. Golgi (potential).  
 CC -!- SIMILARITY: Belongs to the glycosyltransferase family 15.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation  
 at the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; X99619; CAA67930.1; -.  
 CC InterPro; IPR002685; Glyco trans 15.  
 CC Pfam; PF01793; Glyco trans 15; 1.  
 CC Transferrase; Glycosyltransferase; Transmembrane; Signal-anchor;  
 KW Golgi stack.  
 FT DOMAIN 1 9 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 10 28 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)  
 FT (POTENTIAL).  
 FT DOMAIN 29 431 LUMENAL (POTENTIAL).  
 FT ACT\_SITE 318 318 NUCLEOPHILE (POTENTIAL).  
 FT SEQUENCE 431 AA; 50004 MW; E8AC1FAE9010B763 CRC64;  
 SQ  
 Query Match 90.6%; Score 29; DB 1; Length 431;  
 Best Local Similarity 83.3%; Pred. No. 73;  
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 EKAGGF 6  
 DB 338 DKAGGF 343  
 :|||||  
 RESULT 13  
 KTR5 YEAST STANDARD; PRT; 522 AA.  
 ID KTR5 YEAST  
 AC P53966;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Probable mannosyltransferase KTR5 (EC 2.4.1.131).  
 GN KTR5 OR YNL029C OR N2755.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
 OX NCBI\_TaxID=4932;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Dueterhoeft A.; Floeth M.; Fritz C.; Heuss-Neitzel D.;  
 RA Hilbert H.; Moestl D.;  
 RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Andre B.; Iraqui Houssaini I.; Urrestarazu L.A.; Vissers S.;



Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.

[3]

CHARACTERIZATION.

MEDLINE=97245297; PubMed=9090056;

Lusser M., Sdicu A.-M., Winnett E., Vo D.H., Sheraton J., Duesterhoeft A., Storms R.K., Bussey H.;

"Completion of the Saccharomyces cerevisiae genome sequence allows identification of KTR5, KTR6 and KTR7 and definition of the nine-membered KRE2/WNT1 mannosyltransferase gene family in this organism."

Yeast 13:267-274 (1997).

RL CC -!- FUNCTION: POSSIBLE GLYCOSYLTRANSFERASE.

CC -!- CATALYTIC ACTIVITY: Transfers an alpha-D-mannosyl residue from

CC GDP-mannose into lipid-linked oligosaccharide, forming an alpha-

CC 1,2-D-mannosyl-D-mannose linkage.

CC -!- SUBCELLULAR LOCATION: type II membrane protein (Probable).

CC -!- SIMILARITY: Belongs to the glycosyltransferase family 15.

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC

EMBL; Z71305; CAA95891.1; -

DR PIR; S62941; S62941.

DR Germonline; 143036; -

DR SGD; S0004974; KTR5.

DR GO; GO:0007047; P:cell wall organization and biogenesis; IMP.

DR InterPro; IPR002685; Glyco trans 15.

DR Pfam; PF01793; Glyco trans 15; 1.

KW Transferase; Glycosyltransferase; Transmembrane; Signal-anchor;

KW Glycoprotein.

FT DOMAIN 1 16 CYTOPLASMIC (POTENTIAL).

FT TRANSMEM 17 37 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)

FT (POTENTIAL).

FT LUMENAL (POTENTIAL).

FT DOMAIN 38 522 STEM REGION (BY SIMILARITY).

FT DOMAIN 38 82 CATALYTIC (BY SIMILARITY).

FT DOMAIN 83 522 NUCLEOPHILE (POTENTIAL).

FT ACT SITE 363 363 N-LINKED (GLCNAC...) (POTENTIAL).

FT CARBOHYD 86 86

SQ SEQUENCE 522 AA; 61727 MW; 3P9D75D3F09D4110 CRC64;

Query Match 90.6%; Score 29; DB 1; Length 522;

Best Local Similarity 83.3%; Pred. No. 87;

Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 EKAGGF 6

Db 383 EKSGGF 388

RESULT 14

VG42 HAEIN STANDARD; PRT; 631 AA.

ID P44236;

DT 01-NOV-1995 (Rel. 32, Created)

DT 01-NOV-1995 (Rel. 32, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Mu-like prophage FluMu protein gp42.

GN H11514.

OS Haemophilus influenzae.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;

OC Pasteurellaceae; Haemophilus.

OX NCBI\_TaxID=727;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=Rd / KW20 / ATCC 51907;

RX MEDLINE=95350630; PubMed=7542800;

RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F., Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,

McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D., Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M., Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D., Uterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C., Fine L.D., Frichman J.L., Fuhrmann J.U., Geoghagen N.S.M., Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O., Venter J.C.;

"Whole-genome random sequencing and assembly of Haemophilus influenzae Rd."

RL Science 269:496-512 (1995).

RL CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).

CC -!- SIMILARITY: TO PHAGE MU PROTEIN GP42.

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC

EMBL; U32827; AAC23167.1; -

DR PIR; C64034; C64034.

DR TIGR; H11514; -

KW Hypothetical protein; Transmembrane; Complete proteome.

FT TRANSMEM 56 76 POTENTIAL.

FT TRANSMEM 385 405 POTENTIAL.

FT TRANSMEM 455 475 POTENTIAL.

FT TRANSMEM 495 515 POTENTIAL.

FT TRANSMEM 543 563 POTENTIAL.

SQ SEQUENCE 631 AA; 66208 MW; 34E234C986406DE2 CRC64;

Query Match 90.6%; Score 29; DB 1; Length 631;

Best Local Similarity 83.3%; Pred. No. 1e+02;

Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 EKAGGF 6

Db 330 EKAGGY 335

RESULT 15

CAL3 RAT

ID CAL3 RAT STANDARD; PRT; 636 AA.

AC P13941; O70604;

DT 01-JAN-1990 (Rel. 13, Created)

DT 01-JUN-1994 (Rel. 29, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Collagen alpha 1(III) chain (Fragment).

GN COL3A1.

OS Rattus norvegicus (Rat).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

OX NCBI\_TaxID=10116;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=94114571; PubMed=8286415;

RA Glumoff V., Maekelae J.K., Vuorio E.;

RT "Cloning of cDNA for rat pro alpha 1(III) collagen mRNA. Different

RT expression patterns of type I and type III collagen and fibronectin

RT genes in experimental granulation tissue."

RL Biochim. Biophys. Acta 1217:41-48 (1994).

RN [2]

RP SEQUENCE OF 73-636 FROM N.A.

RC STRAIN=Sprague-Dawley; TISSUE=Fibroblast;

RA Wurtz T., Ellerstrom C., Lundmark C., Christersson C.;

RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.

RN [3]

RP SEQUENCE OF 308-482 FROM N.A.

RX MEDLINE=88296083; PubMed=2456904;

RA Frankel F.R., Hsu C.-Y.J., Meyers J.C., Lin E., Lyttle C.R., Komm B., Mohn K.;

RA "Regulation of alpha 2(I), alpha 1(III), and alpha 2(V) collagen

```
RT mRNAs by estradiol in the immature rat uterus." ;
RL DNA 7:347-354(1988).
CC -!- FUNCTION: Collagen type III occurs in most soft connective tissues
CC along with type I collagen.
CC -!- SUBUNIT: Trimers of identical alpha 1(III) chains. The chains are
CC linked to each other by interchain disulfide bonds. Trimers are
CC also cross-linked via hydroxylsines.
CC -!- PTM: Prolines at the third position of the tripeptide repeating
CC unit (G-X-Y) are hydroxylated in some or all of the chains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X70369; CAA49832.1; -.
DR EMBL; AJ005395; CAA06510.1; -.
DR EMBL; M21354; AAA40942.1; -.
DR PIR; S41067; S41067.
DR InterPro; IPR008161; Clq_helix.
DR InterPro; IPR008180; Collagen.
DR InterPro; IPR000885; Fib_collagen_C.
DR InterPro; IPR002181; Fibrinogen_C.
DR InterPro; IPR001007; VWF_C.
DR Pfam; PF01410; COLFI; 1.
DR Pfam; PF01391; Collagen; 6.
DR ProDom; PD000007; Clq_helix; 1.
DR ProDom; PD002078; Fib_collagen_C; 1.
DR SMART; SM00038; COLFI; 1.
DR PROSITE; PS01208; VWFC.1; PARTIAL.
DR Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW Collagen; Glycoprotein.
FT NON_TER 1
FT CHAIN <1 375 COLLAGEN ALPHA 1(III) CHAIN.
FT PROPEP 376 636 CARBOXYL-TERMINAL PROPEPTIDE.
FT DOMAIN <1 368 TRIPLE-HELICAL REGION.
FT DOMAIN 369 636 NONHELICAL REGION (C-TERMINAL).
FT DISULFID 368 368 INTERCHAIN (BY SIMILARITY).
FT DISULFID 369 369 INTERCHAIN (BY SIMILARITY).
FT CONFLICT 340 340 N -> D (IN REF. 2).
FT CONFLICT 429 429 A -> G (IN REF. 2).
SQ SEQUENCE 636 AA; 62332 MW; 61A48159F01D01EE CRC64;

Query Match 90.6%; Score 29; DB 1; Length 636;
Best Local Similarity 83.3%; Pred. No. 1.1e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 EKAGGF 6
Db 381 EKSGGF 386

Search completed: September 18, 2004, 04:26:19
Job time : 6.47059 secs
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 18, 2004, 04:00:08 ; Search time 21.2941 Seconds  
(without alignments)  
88.903 Million cell updates/sec

Title: US-10-615-959-39  
Perfect score: 32  
Sequence: 1 EKAGGF 6

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_virus:\*  
16: sp\_bacteriap:\*  
17: sp\_archaea:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	32	100.0	182	16 Q7VLT3	Q7VLT3 prochloroco
2	32	100.0	198	16 Q81LQ9	Q81LQ9 bacillus an
3	32	100.0	225	16 Q8DV80	Q8DV80 streptococc
4	32	100.0	279	16 Q7VGH7	Q7VGH7 helicobacte
5	32	100.0	805	16 Q888D3	Q888D3 pseudomonas
6	32	100.0	1163	4 Q8N6U4	Q8N6U4 homo sapien
7	32	100.0	1289	16 Q9X087	Q9X087 thermotoga
8	32	100.0	1687	5 Q8T9L7	Q8T9L7 toxoplasma
9	29	90.6	46	16 Q8G360	Q8G360 bruceella su
10	29	90.6	101	10 Q9LGG6	Q9LGG6 oryza sativ
11	29	90.6	127	16 Q8ABW1	Q8ABW1 bacteroides
12	29	90.6	152	2 Q8GGP6	Q8GGP6 streptomyce
13	29	90.6	178	10 Q8LES8	Q8LES8 arabisidopsi
14	29	90.6	178	10 Q84K40	Q84K40 arabisidopsi
15	29	90.6	187	16 Q9A9G0	Q9A9G0 caulobacter
16	29	90.6	199	16 Q818D7	Q818D7 bacillus ce

17	29	90.6	301	16 Q81PN4	Q81PN4 bacillus an
18	29	90.6	322	5 Q8SS28	Q8SS28 encephalito
19	29	90.6	324	16 Q9KXV9	Q9KXV9 streptomyce
20	29	90.6	335	10 Q8GUS9	Q8GUS9 arabisidopsi
21	29	90.6	338	11 Q8BJU6	Q8BJU6 mus musculus
22	29	90.6	341	16 Q89MW1	Q89MW1 bradyrhizob
23	29	90.6	342	2 Q7X3X9	Q7X3X9 clavibacter
24	29	90.6	352	10 Q8GUS8	Q8GUS8 arabisidopsi
25	29	90.6	372	16 Q88SU7	Q88SU7 lactobacill
26	29	90.6	379	16 Q92Q09	Q92Q09 rhizobium m
27	29	90.6	387	2 Q9EUT9	Q9EUT9 rhodococcus
28	29	90.6	392	10 Q84JG8	Q84JG8 oryza sativ
29	29	90.6	410	10 Q7XRS4	Q7XRS4 oryza sativ
30	29	90.6	417	4 Q8NI87	Q8NI87 homo sapien
31	29	90.6	421	10 Q9SIK6	Q9SIK6 arabisidopsi
32	29	90.6	454	9 Q80281	Q80281 bacterioph
33	29	90.6	457	16 Q9ABR3	Q9ABR3 caulobacter
34	29	90.6	490	16 Q8FT53	Q8FT53 corynebacte
35	29	90.6	501	16 Q7UN28	Q7UN28 rhodospirill
36	29	90.6	522	16 Q69811	Q69811 streptomyce
37	29	90.6	537	16 Q92JS7	Q92JS7 rhizobium m
38	29	90.6	555	16 Q8EVD1	Q8EVD1 mycoplasma
39	29	90.6	556	16 Q7UR12	Q7UR12 rhodospirill
40	29	90.6	731	16 Q8XJL1	Q8XJL1 clostridium
41	29	90.6	847	16 Q825F3	Q825F3 streptomyce
42	29	90.6	1007	5 Q96848	Q96848 dictyosteli
43	29	90.6	1007	5 Q23897	Q23897 dictyosteli
44	29	90.6	1075	10 Q65567	Q65567 arabisidopsi
45	29	90.6	1129	16 Q82N96	Q82N96 streptomyce

## ALIGNMENTS

RESULT 1  
Q7VLT3 PRELIMINARY; PRT; 182 AA.  
AC Q7VLT3  
DT 01-OCT-2003 (TREMREL. 25, Created)  
DT 01-OCT-2003 (TREMREL. 25, Last sequence update)  
DT 01-OCT-2003 (TREMREL. 25, Last annotation update)  
DE Hypothetical protein.  
GN PM0772.  
OS Prochlorococcus marinus subsp. pastoris (strain CCMP 1378 / MED4).  
OC Bacteria; Cyanobacteria; Prochlorophytes; Prochlorococcaceae;  
OC Prochlorococcus.  
OX NCBI\_TaxID=59919;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=22825698; PubMed=12917642;  
RA Rocap G., Larimer F.W., Lamerdin J., Malfatti S., Chain P.,  
RA Ahlgren N.A., Arellano A., Coleman M., Hauser L., Hess W.R.,  
RA Johnson Z.I., Land M., Lindell D., Post A.F., Regala W., Shah M.,  
RA Shaw S.L., Steglich C., Sullivan M.B., Ting C.S., Tolonen A.,  
RA Webb E.A., Zinser E.R., Chisholm S.W.;  
RT "Genome divergence in two Prochlorococcus ecotypes reflects oceanic  
niche differentiation."  
RL Nature 424:1042-1047(2003).  
DR EMBL; BX572092; CAB19231.1; --  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 182 AA; 20358 MW; 24C6151DC9065CE6 CRC64;

Query Match 100.0%; Score 32; DB 16; Length 182;  
Best Local Similarity 100.0%; Pred. No. 81;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 EKAGGF 6

Db 47 EKAGGF 52

RESULT 2  
Q81LQ9

Q8L1Q9 PRELIMINARY; PRT; 198 AA.

Q8L1Q9; 01-JUN-2003 (TREMBlrel. 24, Created)

DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)

DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)

DE Come operon protein 1.

GN COMEA OR BA4553.

OS Bacillus anthracis (strain Ames).

OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.

OX NCBI\_TaxID=198094;

RN [1]

RX MEDLINE=22608414; PubMed=12721629;

RA Read T.D., Peterson S.N., Tourasse N., Baillie L.W., Paulsen I.T., Nelson K.E., Tetzelin H., Fouts D.E., Eisen J.A., Gill S.R., Holtzapfe E.K., Okstad O.A., Helgason E., Rilstone J., Wu M., Kolonay J.F., Beanan W.J., Dodson R.C., Durkin A.S., Haft D.H., DeBoy R.T., Madpu R., Daugherty S.C., Hance I.R., Weidman J.F., Nelson W.C., Peterson J.D., Pop M., Khouri H.M., Radune D., Benton J.L., Mahamoud Y., Jiang L., Hance I.R., Nierman W.C., Berry K.J., Plaut R.D., Wolf A.M., Watkins K.L., Nierman W.C., Hazen A., Cline R., Redmond C., Thwaite J.E., White O., Salzberg S.L., Thomason B., Friedlander A.M., Koehler T.M., Hanna P.C., Kolsto A.-B., Fraser C.M.;

RT "The genome sequence of Bacillus anthracis Ames and comparison to RT closely related bacteria.";

RL Nature 423:81-86(2003).

DR EMBL; AE017038; AAP28261.1; --

DR TIGR; BA4553; --

DR GO: 0005622; C:intracellular; IEA.

DR GO: 0003677; F:DNA binding; IEA.

DR InterPro; IPR004509; COMEA\_HHH.

DR InterPro; IPR000445; Hhh\_1.

DR SMART; PF00633; HHH; 2.

DR SMART; SM00278; Hhh1; 2.

DR TIGRFAMs; TIGR00426; TIGR00426; 1.

SQ SEQUENCE 198 AA; 22368 MW; 8A8C008985CC3E8 CRC64;

Query Match 100.0%; Score 32; DB 16; Length 198;  
Best Local Similarity 100.0%; Pred. No. 89;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 EKAGGF 6  
| | | | |  
Db 89 EKAGGF 94

RESULT 3

Q8DV80 Q8DV80 PRELIMINARY; PRT; 225 AA.

AC Q8DV80;

DT 01-MAR-2003 (TREMBlrel. 23, Created)

DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)

DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)

DE Putative competence protein.

GN COMEA OR SMU.625.

OS Streptococcus mutans.

OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;

OC Streptococcus.

OX NCBI\_TaxID=1309;

RN [1]

RX STRAIN=UA159 / ATCC 700610 / Serotype C;

RA Ajdic D., McShan W.M., McLaughlin R.E., Savic G., Chang J., Carson M.B., Primeaux C., Tian R., Kenton S., Jia H., Lin S., Li S., Zhu H., Najjar F., Lai H., White J., Roe B.A., Ferretti J.J.;

RT "Genome sequence of Streptococcus mutans UA159, a cariogenic dental pathogen.";

RL Proc. Natl. Acad. Sci. U.S.A. 99:14434-14439(2002).

GO: 0005622; C:intracellular; IEA.  
GO: 0003677; F:DNA binding; IEA.  
InterPro; IPR004509; COMEA\_HHH.  
InterPro; IPR000445; Hhh\_1.  
SMART; PF00633; HHH; 2.  
SMART; SM00278; Hhh1; 2.  
TIGRFAMs; TIGR00426; TIGR00426; 1.  
Complete proteome.  
SEQUENCE 198 AA; 22368 MW; 8A8C008985CC3E8 CRC64;

Query Match 100.0%; Score 32; DB 16; Length 198;  
Best Local Similarity 100.0%; Pred. No. 89;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 EKAGGF 6  
| | | | |  
Db 89 EKAGGF 94

RESULT 3

Q8DV80 Q8DV80 PRELIMINARY; PRT; 225 AA.

AC Q8DV80;

DT 01-MAR-2003 (TREMBlrel. 23, Created)

DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)

DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)

DE Putative competence protein.

GN COMEA OR SMU.625.

OS Streptococcus mutans.

OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;

OC Streptococcus.

OX NCBI\_TaxID=1309;

RN [1]

RX STRAIN=UA159 / ATCC 700610 / Serotype C;

RA Ajdic D., McShan W.M., McLaughlin R.E., Savic G., Chang J., Carson M.B., Primeaux C., Tian R., Kenton S., Jia H., Lin S., Li S., Zhu H., Najjar F., Lai H., White J., Roe B.A., Ferretti J.J.;

RT "Genome sequence of Streptococcus mutans UA159, a cariogenic dental pathogen.";

RL Proc. Natl. Acad. Sci. U.S.A. 99:14434-14439(2002).

GO: 0005622; C:intracellular; IEA.  
GO: 0003677; F:DNA binding; IEA.  
InterPro; IPR004509; COMEA\_HHH.  
InterPro; IPR000445; Hhh\_1.  
SMART; PF00633; HHH; 2.  
SMART; SM00278; Hhh1; 2.  
TIGRFAMs; TIGR00426; TIGR00426; 1.  
Complete proteome.  
SEQUENCE 198 AA; 22368 MW; 8A8C008985CC3E8 CRC64;

Query Match 100.0%; Score 32; DB 16; Length 198;  
Best Local Similarity 100.0%; Pred. No. 89;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 EKAGGF 6  
| | | | |  
Db 89 EKAGGF 94

RESULT 3

Q8DV80 Q8DV80 PRELIMINARY; PRT; 225 AA.

AC Q8DV80;

DT 01-MAR-2003 (TREMBlrel. 23, Created)

DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)

DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)

DE Putative competence protein.

GN COMEA OR SMU.625.

OS Streptococcus mutans.

OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;

OC Streptococcus.

OX NCBI\_TaxID=1309;

RN [1]

RX STRAIN=UA159 / ATCC 700610 / Serotype C;

RA Ajdic D., McShan W.M., McLaughlin R.E., Savic G., Chang J., Carson M.B., Primeaux C., Tian R., Kenton S., Jia H., Lin S., Li S., Zhu H., Najjar F., Lai H., White J., Roe B.A., Ferretti J.J.;

RT "Genome sequence of Streptococcus mutans UA159, a cariogenic dental pathogen.";

RL Proc. Natl. Acad. Sci. U.S.A. 99:14434-14439(2002).

GO: 0005622; C:intracellular; IEA.  
GO: 0003677; F:DNA binding; IEA.  
InterPro; IPR004509; COMEA\_HHH.  
InterPro; IPR000445; Hhh\_1.  
SMART; PF00633; HHH; 2.  
SMART; SM00278; Hhh1; 2.  
TIGRFAMs; TIGR00426; TIGR00426; 1.  
Complete proteome.  
SEQUENCE 198 AA; 22368 MW; 8A8C008985CC3E8 CRC64;

Query Match 100.0%; Score 32; DB 16; Length 198;  
Best Local Similarity 100.0%; Pred. No. 89;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 EKAGGF 6  
| | | | |  
Db 89 EKAGGF 94

RESULT 3

Q8DV80 Q8DV80 PRELIMINARY; PRT; 225 AA.

AC Q8DV80;

DT 01-MAR-2003 (TREMBlrel. 23, Created)

DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)

DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)

DE Putative competence protein.

GN COMEA OR SMU.625.

OS Streptococcus mutans.

OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;

OC Streptococcus.

OX NCBI\_TaxID=1309;

RN [1]

RX STRAIN=UA159 / ATCC 700610 / Serotype C;

RA Ajdic D., McShan W.M., McLaughlin R.E., Savic G., Chang J., Carson M.B., Primeaux C., Tian R., Kenton S., Jia H., Lin S., Li S., Zhu H., Najjar F., Lai H., White J., Roe B.A., Ferretti J.J.;

RT "Genome sequence of Streptococcus mutans UA159, a cariogenic dental pathogen.";

RL Proc. Natl. Acad. Sci. U.S.A. 99:14434-14439(2002).

GO: 0005622; C:intracellular; IEA.  
GO: 0003677; F:DNA binding; IEA.  
InterPro; IPR004509; COMEA\_HHH.  
InterPro; IPR000445; Hhh\_1.  
SMART; PF00633; HHH; 2.  
SMART; SM00278; Hhh1; 2.  
TIGRFAMs; TIGR00426; TIGR00426; 1.  
Complete proteome.  
SEQUENCE 198 AA; 22368 MW; 8A8C008985CC3E8 CRC64;

Query Match 100.0%; Score 32; DB 16; Length 198;  
Best Local Similarity 100.0%; Pred. No. 89;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 EKAGGF 6  
| | | | |  
Db 89 EKAGGF 94

RESULT 3

Q8DV80 Q8DV80 PRELIMINARY; PRT; 225 AA.

AC Q8DV80;

DT 01-MAR-2003 (TREMBlrel. 23, Created)

DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)

DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)

DE Putative competence protein.

GN COMEA OR SMU.625.

OS Streptococcus mutans.

OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;

OC Streptococcus.

OX NCBI\_TaxID=1309;

RN [1]

RX STRAIN=UA159 / ATCC 700610 / Serotype C;

RA Ajdic D., McShan W.M., McLaughlin R.E., Savic G., Chang J., Carson M.B., Primeaux C., Tian R., Kenton S., Jia H., Lin S., Li S., Zhu H., Najjar F., Lai H., White J., Roe B.A., Ferretti J.J.;

RT "Genome sequence of Streptococcus mutans UA159, a cariogenic dental pathogen.";

RL Proc. Natl. Acad. Sci. U.S.A. 99:14434-144

```

RA Buell R., Joardar V., Khouri H., Fedorova N., Tran B., Russell D.,
RA Berry K., Utterback T., Van Aken S., Feldblyum T., Gwinn M.,
RA Dodson R., Deboy R., Durkin A., Kolonay J., Madupu R., Daugherty S.,
RA Brinkac L., Beanan M., Haft D., Selengut J., Nelson W., Davidse T.,
RA White O., Fraser C., Collmer A.;
RT "Complete sequence of Pseudomonas syringae.";
RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE016859; AA054622.1; -.
DR TIGR; FSPT01093; -.
DR InterPro; IPR005094; Relaxase.
DR Pfam; PF03432; Relaxase; 1.
KW Complete proteome.
SQ SEQUENCE 805 AA; 92502 MW; 2E8D5D47356FAA14 CRC64;

Query Match 100.0%; Score 32; DB 16; Length 805;
Best Local Similarity 100.0%; Pred. No. 3.7e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EKAGGF 6
DB 676 EKAGGF 681

RESULT 6
Q8NGU4 PRELIMINARY; PRT; 1163 AA.
AC Q8NGU4;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Similar to collagen, type III, alpha 1 (Ehlers-Danlos syndrome type
DE IV, autosomal dominant).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RA Strausberg R.;
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC028178; AAH28178.1; -.
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR002181; Fibrinogen_C.
DR InterPro; IPR000885; Fib.collagen_C.
DR Pfam; PF01410; COLFI; 1.
DR Pfam; PF01391; Collagen; 13.
DR ProDom; PD002078; Fib_collagen_C; 1.
DR SMART; SM00038; COLFI; 1.
DR SMART; SM00214; VWC; 1.
DR PROSITE; PS01208; VWC_1; 1.
DR PROSITE; PS50184; VWC_2; 1.
KW Collagen.
SQ SEQUENCE 1163 AA; 111899 MW; 9E0C6BEE1E94D6357 CRC64;

Query Match 100.0%; Score 32; DB 4; Length 1163;
Best Local Similarity 100.0%; Pred. No. 5.4e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EKAGGF 6
DB 908 EKAGGF 913

RESULT 7
Q9X087 PRELIMINARY; PRT; 1289 AA.
AC Q9X087;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)

```

```

DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein TM0992.
GN TM0992.
OC Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.
OX NCBI_TaxID=2336;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MSB8 / DSM 3109;
RX MEDLINE=99287316; PubMed=10360571;
RA Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J.,
RA Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,
RA McDonald L., Utterback T.R., Malek J.A., Linher K.D., Garrett M.M.,
RA Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,
RA Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,
RA Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;
RT "Evidence for lateral gene transfer between Archaea and Bacteria from
RT genome sequence of Thermotoga maritima.";
RL Nature 399:323-329 (1999).
DR EMBL; AE001761; AAD36071.1; -.
DR PIR; F72308; F72308.
DR TIGR; TM0992; -.
DR InterPro; IPR008947; PLC_Nuclease
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 1289 AA; 148491 MW; 82E5081AD98ADCAA CRC64;

Query Match 100.0%; Score 32; DB 16; Length 1289;
Best Local Similarity 100.0%; Pred. No. 6e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EKAGGF 6
DB 942 EKAGGF 947

RESULT 8
Q8T9L7 PRELIMINARY; PRT; 1687 AA.
AC Q8T9L7;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Carbamoyl phosphate synthetase II (EC 6.3.5.5).
OS Toxoplasma gondii.
OC Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida; Sarcocystidae;
OC Toxoplasma.
OX NCBI_TaxID=5811;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RH;
RX MEDLINE=21848414; PubMed=11859373;
RA Fox B.A., Bzik D.J.;
RT "De novo pyrimidine biosynthesis is required for virulence of
RT Toxoplasma gondii.";
RL Nature 415:926-929 (2002).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=RH;
RA Fox B.A., Bzik D.J.;
RT "Novel organization and sequences of glutamine-dependent carbamoyl
RT phosphate synthetase II in Toxoplasma gondii.";
RL Int. J. Parasitol. 0:0-0 (2002).
DR EMBL; AY059630; AAL27793.2; -.
DR GO; GO:0005737; C:cytoplasm; IEA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0004086; F:carbamoyl-phosphate synthase (glutamine-hyd. . .); IEA.
DR GO; GO:0016874; F:ligase activity; IEA.
DR GO; GO:0006526; P:arginine biosynthesis; IEA.
DR GO; GO:0006807; P:nitrogen metabolism; IEA.
DR GO; GO:0019856; P:pyrimidine base biosynthesis; IEA.
DR InterPro; IPR006220; Anth_synthII.
DR InterPro; IPR006275; CarA_L_glu.
DR InterPro; IPR006274; CarA_synth_small.

```

RESULT 10  
O9LGG6

Db 60 EKSGGF 65

Db 154 ERAGGF 159

## RESULT 12

Q8GGE6 PRELIMINARY; PRT; 152 AA.  
 AC Q8GGE6;  
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
 DE Hypothetical protein.  
 OS Streptomyces sp. EN27.  
 OG Plasmid pEN2701.  
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 OC Streptomycineae; Streptomycetaceae; Streptomycetes.  
 OX NCBI\_TaxID=211464;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=EN27;  
 RA Coombs J.T., Franco C.M.M., Loria R.;  
 RT "Complete sequencing and analysis of pEN2701, a novel 13kb plasmid  
 from an endophytic Streptomyces sp.";  
 RL Plasmid 0:0-0(2002).  
 DR EMBL; AF533985; AAN76280.1; -.  
 DR GO; GO:0046821; C:extrachromosomal DNA; IEA.  
 KW Hypothetical protein; Plasmid.  
 SQ SEQUENCE 152 AA; 15024 MW; 1EB21C8DA9FC9D56 CRC64;

Query Match 90.6%; Score 29; DB 2; Length 152;

Best Local Similarity 83.3%; Pred. No. 2.9e+02;

Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 EKAGGF 6

Db 50 DKAGGF 55

## RESULT 13

Q8LES8 PRELIMINARY; PRT; 178 AA.  
 AC Q8LES8;  
 DT 01-OCT-2002 (TrEMBLrel. 22, Created)  
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)  
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)  
 DE Hypothetical protein.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Haas B.J., Volfovsky N., Town C.D., Troughan M., Alexandrov N.,  
 RA Feldmann K.A., Flavell R.B., White O., Salzberg S.L.;  
 RT "Full-length messenger RNA sequences greatly improve genome  
 annotation.";  
 RL Genome Biol. 0:0-0(2002).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Brover V., Troughan M., Alexandrov N., Lu Y.-P., Flavell R.,  
 RA Feldmann K.;  
 RT "Full-length cDNA from Arabidopsis thaliana";  
 RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AY085254; AAM62486.1; -.  
 KW Hypothetical protein.  
 SQ SEQUENCE 178 AA; 20047 MW; 3A0D3BBA314EF24 CRC64;

Query Match 90.6%; Score 29; DB 10; Length 178;

Best Local Similarity 83.3%; Pred. No. 3.4e+02;

Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 EKAGGF 6

Db 154 ERAGGF 159

## RESULT 14

Q84K40 PRELIMINARY; PRT; 178 AA.  
 AC Q84K40;  
 DT 01-JUN-2003 (TrEMBLrel. 24, Created)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
 DE Hypothetical protein.  
 GN AT4G30820.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Yamada K., Chan M.M., Chang C.H., Dale J.M., Heuan V.W., Lee J.M.,  
 RA Onodera C.S., Quach H.L., Tang C., Toriumi M., Wong C., Wu H.C.,  
 RA Yu G., Yuan S., Carninci P., Chen H., Cheuk R., Hayashizaki Y.,  
 RA Ishida J., Jones T., Kamiya A., Kawai J., Kim C.J., Narusaka M.,  
 RA Nguyen M., Palm C.J., Sakurai T., Satou M., Seki M., Shim P.,  
 RA Southwick A., Tripp M.G., Wu T., Shinozaki K., Davis R.W., Ecker J.R.,  
 RA Theologis A.;  
 RT "Arabidopsis Full Length cDNA Clones";  
 RL Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Yamada K., Chan M.M., Chang C.H., Dale J.M., Heuan V.W., Lee J.M.,  
 RA Onodera C.S., Quach H.L., Tang C., Toriumi M., Wong C., Wu H.C.,  
 RA Yu G., Yuan S., Carninci P., Chen H., Cheuk R., Hayashizaki Y.,  
 RA Ishida J., Jones T., Kamiya A., Kawai J., Kim C.J., Narusaka M.,  
 RA Nguyen M., Palm C.J., Sakurai T., Satou M., Seki M., Shim P.,  
 RA Southwick A., Tripp M.G., Wu T., Shinozaki K., Davis R.W., Ecker J.R.,  
 RA Theologis A.;  
 RT "Arabidopsis Full Length cDNA Clones";  
 RL Submitted (FEB-2003) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; BT002819; AAO22638.1; -.  
 DR EMBL; BT004364; AAO42358.1; -.  
 KW Hypothetical protein.  
 SQ SEQUENCE 178 AA; 20077 MW; 28614BA6CF64EF24 CRC64;

Query Match 90.6%; Score 29; DB 10; Length 178;

Best Local Similarity 83.3%; Pred. No. 3.4e+02;

Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 EKAGGF 6

Db 154 ERAGGF 159

## RESULT 15

Q9A9G0 PRELIMINARY; PRT; 187 AA.  
 AC Q9A9G0;  
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
 DE Hypothetical protein CC1028.  
 GN CC1028.  
 OS Caulobacter crescentus.  
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Caulobacteriales;  
 OC Caulobacteraceae; Caulobacter.  
 OX NCBI\_TaxID=155892;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 19089 / CB15;  
 RX MEDLINE=21173698; PubMed=11259647;  
 RA Nierman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,  
 RA Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,  
 RA Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,

Mon Sep 20 11:05:42 2004

RA DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,  
RA Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,  
RA Utterback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,  
RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;  
RT "Complete genome sequence of *Caulobacter crescentus*.";   
RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).  
DR EMBL; AE005780; AAK23012.1; -.  
DR PIR; H87376; H87376.  
DR TIGR; CC1028; -.  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 187 AA; 19496 MW; 9C2B20F466B2D566 CRC64;  
Query Match 90.8%; Score 29; DB 16; Length 187;  
Best Local Similarity 83.3%; Pred. No. 3.6e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
QY 1 ERAGGF 6  
Db 126 ERAGGF 131

Search completed: September 18, 2004, 04:29:34  
Job time : 25.2941 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 18, 2004, 03:47:39 ; Search time 42.8235 Seconds  
(without alignments)  
52.784 Million cell updates/sec

Title: US-10-615-959-40

Perfect score: 42

Sequence: 1 IGGEXAGG 8

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_29Jan04:\*

1: geneseqp1980s:\*

2: geneseqp1990s:\*

3: geneseqp2000s:\*

4: geneseqp2001s:\*

5: geneseqp2002s:\*

6: geneseqp2003as:\*

7: geneseqp2003bs:\*

8: geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	42	100.0	8	AAB61740	Aab61740 Human typ
2	42	100.0	9	AAB61741	Aab61741 Human typ
3	42	100.0	11	AAB61742	Aab61742 Human typ
4	42	100.0	12	AAB61743	Aab61743 Human typ
5	42	100.0	25	AAB38632	Aae38632 Human C-t
6	42	100.0	30	AAR92068	Aar92068 Recombina
7	42	100.0	262	AAR38640	Aae38640 Human col
8	42	100.0	270	AAY07375	Aay07375 Procollag
9	42	100.0	293	ABU70813	Abu70813 Human adi
10	42	100.0	309	ABU23675	Aau23675 Novel hum
11	42	100.0	309	ABG60248	Abg60248 Human ova
12	42	100.0	309	ABG61719	Abg61719 Novel ova
13	42	100.0	339	ABU70735	Abu70735 Human ova
14	42	100.0	623	AAW12843	AAw12843 Human adi
15	42	100.0	626	AAW12842	AAw12842 Pro-alpha
16	42	100.0	842	ADS08475	Ades08475 Novel pro
17	42	100.0	1466	ABB50291	Abb50291 Collagen
18	42	100.0	1466	ABB90747	Abb90747 Human tum
19	42	100.0	1466	ABU54454	Abu54454 Human tum
20	42	100.0	1466	ABR47418	ABr47418 Breast ca
21	42	100.0	1469	ABG15191	ABg15191 Novel hum
22	42	100.0	1470	ABU09399	ABu09399 Novel pro
23	42	100.0	1726	ABR42661	ABr42661 Decorin-m
24	41	97.6	1466	AAE02537	AAe02537 Porcine a
25	38	90.5	142	AAg40971	AAg40971 Zea mays

26	38	90.5	674	2	AAy29125	Aay29125 Amino aci
27	37	88.1	481	6	ABM70142	Abm70142 Photorhab
28	36	85.7	11	3	AAy68425	Aay68425 Type III
29	36	85.7	11	3	AAy68436	Aay68436 Type III
30	36	85.7	11	3	AAy82167	Aay82167 Type III
31	36	85.7	11	4	AAy82167	Aay82167 Type II a
32	36	85.7	1078	4	AAy82167	Aay82167 Collagen
33	36	85.7	1078	3	AAy96125	Aay96125 Collagen
34	36	85.7	1078	5	AAE16478	Aae16478 Human col
35	36	85.7	1078	5	ABB80736	Abb80736 Collagen
36	36	85.7	1078	5	ABB09628	Abb09628 Amino aci
37	35	83.3	128	2	AAy20721	Aay20721 Human neu
38	35	83.3	162	3	AAG34474	Aag34474 Arabidops
39	35	83.3	320	6	ABP70959	Abp70959 Epoxide h
40	35	83.3	357	4	ABG08457	Abg08457 Novel hum
41	35	83.3	601	4	ABB64897	Abb64897 Drosophil
42	35	83.3	691	6	ABU19644	Abu19644 Protein e
43	35	83.3	1464	2	AAR88469	Aar88469 Feline in
44	35	83.3	1466	4	AAE02534	Aae02534 Bovine al
45	35	83.3	1466	4	AAE02533	Aae02533 Bovine al

## ALIGNMENTS

RESULT 1  
AAB61740  
ID AAB61740 standard; peptide; 8 AA.  
XX  
AC AAB61740;  
XX  
XX  
DT 20-APR-2001 (first entry)  
XX  
DE Human type III collagen carboxy-telopeptide fragment.  
XX  
XX  
KW Type II collagen; immunoassay; cartilage; telopeptide; human;  
KW type III collagen.  
XX  
OS Synthetic.  
OS Homo sapiens.  
XX  
PN WO200079284-A1.  
XX  
PD 28-DEC-2000.  
XX  
PF 10-DEC-1999; 99WO-US029357.  
XX  
PR 17-JUN-1999; 99US-00335098.  
PR 29-JUN-1999; 99US-0141574P.  
PR 02-JUL-1999; 99US-0142274P.  
PR 07-JUL-1999; 99US-0142675P.  
PR 30-AUG-1999; 99US-00385740.  
XX  
PA (WASH-) WASHINGTON RES FOUND.  
XX  
PI Eyre DR;  
XX  
XX WPI; 2001-146859/15.  
XX  
PT Assay for detecting cross-linked telopeptide analytes indicative of type  
PT II collagen resorption in vivo in a body fluid sample, comprises  
PT contacting the sample with an antibody which binds to the analyte.  
XX  
PS Disclosure; Page 15; 34pp; English.  
XX  
CC The invention relates to immunoassays for measuring type II collagen  
CC (cartilage) resorption in vivo. The method of analysing a body fluid  
CC sample for the presence of an analyte indicative of a physiological  
CC condition, involves contacting the body fluid sample with an antibody  
CC (Ab) which binds to the analyte, detecting binding of Ab in the body  
CC fluid sample, and correlating any detected binding to the physiological  
CC condition. The analysis is useful for measuring type II collagen  
CC (cartilage) resorption in vivo, for distinguishing between resorption of

CC non-mineralized and mineralized cartilage, and for measuring total  
CC cartilage resorption in vivo. The present sequence represents a linear  
CC synthetic peptide based on the carboxy-terminal telopeptide sequence of  
CC human collagen type III  
XX  
SQ Sequence 8 AA;

Query Match 100.0%; Score 42; DB 4; Length 8;  
Best Local Similarity 100.0%; Pred. No. 1.4e+06;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 IGGEKAGG 8  
| | | | |  
Db 1 IGGEKAGG 8

RESULT 2  
AAB61741  
ID AAB61741 standard; peptide; 9 AA.  
XX  
AC AAB61741;  
XX  
DT 20-APR-2001 (first entry)  
XX  
DE Human type III collagen carboxy-telopeptide fragment.  
XX  
KW Type II collagen; immunoassay; cartilage; telopeptide; human;  
XX  
OS Synthetic.  
OS Homo sapiens.  
XX  
PN WO200079284-A1.  
XX  
PD 28-DEC-2000.  
XX  
PF 10-DEC-1999; 99WO-US029357.  
XX  
PR 17-JUN-1999; 99US-00335098.  
PR 29-JUN-1999; 99US-0141574P.  
PR 02-JUL-1999; 99US-0142274P.  
PR 07-JUL-1999; 99US-0142675P.  
PR 30-AUG-1999; 99US-00385740.  
XX  
PA (WASH-) WASHINGTON RES FOUND.  
XX  
PI Eyre DR;  
XX  
DR WPI; 2001-146859/15.  
XX  
PT Assay for detecting cross-linked telopeptide analytes indicative of type  
PT II collagen resorption in vivo in a body fluid sample, comprises  
PT contacting the sample with an antibody which binds to the analyte.  
XX  
PS Disclosure; Page 15; 34pp; English.  
XX  
CC The invention relates to immunoassays for measuring type II collagen  
CC (cartilage) resorption in vivo. The method of analysing a body fluid  
CC sample for the presence of an analyte indicative of a physiological  
CC condition, involves contacting the body fluid sample with an antibody  
CC (Ab) which binds to the analyte, detecting binding of Ab in the body  
CC fluid sample, and correlating any detected binding to the physiological  
CC condition. The analysis is useful for measuring type II collagen  
CC (cartilage) resorption in vivo, for distinguishing between resorption of  
CC non-mineralized and mineralized cartilage, and for measuring total  
CC cartilage resorption in vivo. The present sequence represents a linear  
CC synthetic peptide based on the carboxy-terminal telopeptide sequence of  
CC human collagen type III  
XX  
SQ Sequence 9 AA;

Query Match 100.0%; Score 42; DB 4; Length 9;  
Best Local Similarity 100.0%; Pred. No. 1.4e+06;

Qy 1 IGGEKAGG 8  
| | | | |  
Db 4 IGGEKAGG 11

RESULT 4  
AAB61743

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 IGGEKAGG 8  
| | | | |  
Db 4 IGGEKAGG 11

Query Match 100.0%; Score 42; DB 4; Length 11;  
Best Local Similarity 100.0%; Pred. No. 0.48;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 IGGEKAGG 8  
| | | | |  
Db 4 IGGEKAGG 11

RESULT 4  
AAB61743

ID AAB61743 standard; peptide; 12 AA.  
 AC AAB61743;  
 XX 20-APR-2001 (first entry)  
 DT Human type III collagen carboxy-telopeptide fragment.  
 DE XX  
 DE XX  
 DE XX  
 KW Type II collagen; immunoassay; cartilage; telopeptide; human;  
 KW type III collagen.  
 XX  
 XX Synthetic.  
 OS Homo sapiens.  
 OS WO200079284-A1.  
 XX  
 XX 28-DEC-2000.  
 PD  
 PD  
 PD 10-DEC-1999; 99WO-US029357.  
 PF  
 PF 17-JUN-1999; 99US-00335098.  
 PR 29-JUN-1999; 99US-0141574P.  
 PR 02-JUL-1999; 99US-0142274P.  
 PR 07-JUL-1999; 99US-0142675P.  
 PR 30-AUG-1999; 99US-00385740.  
 XX  
 XX (WASH-) WASHINGTON RES FOUND.  
 PA  
 PA Eyre DR;  
 XX  
 XX WPI; 2001-146859/15.  
 DR  
 DR Assay for detecting cross-linked telopeptide analytes indicative of type  
 PT II collagen resorption in vivo in a body fluid sample, comprises  
 PT contacting the sample with an antibody which binds to the analyte.  
 PT  
 XX Disclosure; Page 15; 34pp; English.  
 PS  
 PS The invention relates to immunoassays for measuring type II collagen  
 CC (cartilage) resorption in vivo. The method of analysing a body fluid  
 CC sample for the presence of an analyte indicative of a physiological  
 CC condition, involves contacting the body fluid sample with an antibody  
 CC (Ab) which binds to the analyte, detecting binding of Ab in the body  
 CC fluid sample, and correlating any detected binding to the physiological  
 CC condition. The analysis is useful for measuring type II collagen  
 CC (cartilage) resorption in vivo, for distinguishing between resorption of  
 CC non-mineralized and mineralized cartilage, and for measuring total  
 CC cartilage resorption in vivo. The present sequence represents a linear  
 CC synthetic peptide based on the carboxy-terminal telopeptide sequence of  
 CC human collagen type III  
 XX  
 SQ Sequence 12 AA;  
 Query Match 100.0%; Score 42; DB 4; Length 12;  
 Best Local Similarity 100.0%; Pred. No. 0.53;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 IGGEKAGG 8  
 DB 4 IGGEKAGG 11  
 |||||  
 RESULT 5  
 AAE38632  
 ID AAE38632 standard; peptide; 25 AA.  
 XX  
 XX AAE38632;  
 AC  
 XX 04-DEC-2003 (first entry)  
 DT  
 DT Human C-terminal telopeptide alpha 1 #4.  
 DE  
 DE Human; therapy; osteoporosis; left-ventricular hypertrophy; scleroderma;  
 XX  
 KW

KW liver cirrhosis; congestive heart failure; multiple sclerosis; fibrosis;  
 KW coal-miner's pneumoconiosis; muscle-wasting syndrome; diabetes mellitus;  
 KW rheumatoid arthritis; Alzheimer's disease; muscular dystrophy; cancer;  
 KW telopeptide alpha.  
 XX Homo sapiens.  
 OS  
 XX WO2003068919-A2.  
 FN 21-AUG-2003.  
 PD  
 PD 12-FEB-2003; 2003WO-US004183.  
 PF  
 PF 12-FEB-2002; 2002US-0356008P.  
 PR  
 PR (REGC ) UNIV CALIFORNIA.  
 XX  
 XX Hellerstein MK;  
 XX  
 XX WPI; 2003-689661/65.  
 DR  
 DR Determining rate of biosynthesis or breakdown of inaccessible biological  
 PT molecules, useful e.g. for diagnosis or monitoring treatment, by  
 PT administering labeled precursor.  
 PT  
 XX Claim 24; Page 99; 105pp; English.  
 PS  
 PS The invention relates to a method of determining the rate of biosynthesis  
 CC or breakdown of at least one inaccessible biological molecule in a  
 CC subject. The method is useful for diagnosis or monitoring and treatment  
 CC of diseases associated with an altered rate of biosynthesis/breakdown of  
 CC an isotopically labelled precursor molecule, specifically osteoporosis;  
 CC left-ventricular hypertrophy; liver cirrhosis or fibrosis; congestive  
 CC heart failure; scleroderma; coal-miner's pneumoconiosis; cardiac or lung  
 CC fibrosis; Alzheimer's disease; multiple sclerosis; rheumatoid arthritis;  
 CC diabetes mellitus; muscle-wasting syndromes; muscular dystrophy; athletic  
 CC training and cancer. The method is also useful for screening candidate  
 CC gene or protein targets, phenotypic/human validation studies on potential  
 CC drugs, drug mechanism studies and determining the risk of developing the  
 CC disease. The present sequence is human C-terminal telopeptide alpha 1.  
 CC This sequence is used to illustrate the method of the invention  
 XX  
 SQ Sequence 25 AA;  
 Query Match 100.0%; Score 42; DB 7; Length 25;  
 Best Local Similarity 100.0%; Pred. No. 1.2;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 IGGEKAGG 8  
 DB 12 IGGEKAGG 19  
 |||||  
 RESULT 6  
 AAR92068  
 ID AAR92068 standard; peptide; 30 AA.  
 XX  
 XX AAR92068;  
 AC  
 XX 27-SEP-1996 (first entry)  
 DT  
 DT Recombinant alpha(III)collagen C-terminus.  
 XX  
 XX Alpha(III)collagen; human; pro-collagen; pro-peptide; artificial skin;  
 KW proteolytic cleavage site; tissue; biocompatible material; cell culture;  
 KW suture; haemostatic sponge; tissue augmentation.  
 XX Homo sapiens.  
 OS  
 XX Key Location/Qualifiers  
 FH Misc-difference 16  
 FT Misc-difference 17  
 FT /note= "any amino acid except Asp"

FT /note= "any amino acid except Glu"  
FT Misc-difference 18  
FT /note= "any amino acid except Pro"  
FT Misc-difference 19  
FT /note= "any amino acid except Met"  
FT Misc-difference 20  
FT /note= "any amino acid except Asp"  
FT Misc-difference 21  
FT /note= "any amino acid except Phe"  
FT Misc-difference 22  
FT /note= "any amino acid except Lys"  
FT Misc-difference 23  
FT /note= "any amino acid except Ile"  
FT Misc-difference 24  
FT /note= "any amino acid except Asn"  
FT Misc-difference 25  
FT /note= "any amino acid except Thr"  
FT Misc-difference 26  
FT /note= "any amino acid except Asp"  
FT Misc-difference 27  
FT /note= "any amino acid except Glu"  
FT Misc-difference 28  
FT /note= "any amino acid except Ile"  
FT Misc-difference 29  
FT /note= "any amino acid except Met"  
FT Misc-difference 30  
FT /note= "any amino acid except Thr"  
FT  
FT  
PN EP699752-A2.  
XX  
XX  
XX  
PD 06-MAR-1996.  
XX  
XX  
XX 30-MAY-1995; 95EP-00108307.  
XX  
XX 22-JUL-1994; 94US-00278774.  
XX  
XX (CLGE ) COLLAGEN CORP.  
XX  
XX Berg RA, Toman PD, Wallace DG;  
XX WPI; 1996-130769/14.  
XX  
XX Recombinant production of collagen - by expressing a pro-peptide-collagen  
XX sequence and cleaving at an intermediate proteolytic recognition site.  
XX  
XX Disclosure; Page 5; 27pp; English.

XX AAR92061-R92069 represent N-terminal and C-terminal peptides of the  
XX recombinant human collagen polypeptides of the invention. This sequence  
XX represents the C-terminal peptide of a recombinant human  
XX alpha1(III)collagen of the invention. The recombinant pro-collagen of the  
XX invention comprises a natural collagen polypeptide chain, a pro-peptide,  
XX and a non-natural site-specific proteolytic agent recognition site  
XX between the collagen and pro-peptide. The recombinant pro-collagens are  
XX used to produce collagens which can be used in tissue and cell cultures.  
XX The collagens can also be used as biocompatible materials such as  
XX artificial skin, sutures, haemostatic sponges or tissue augmentation  
XX compositions for use in humans. The pro-peptide increases the yield of  
XX secreted pro-collagen from cells expressing the recombinant pro-collagen.  
XX The increase in yield of the pro-collagen, as compared to cells  
XX expressing the collagen chains alone, is at least 1000%

XX Sequence 30 AA;

Query Match 100.0%; Score 42; DB 2; Length 30;  
Best Local Similarity 100.0%; Pred. No. 1.4;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 IGGEXAGG 8  
Db 2 IGGEXAGG 9  
|||||

RESULT 7  
AAE38640  
ID AAE38640 standard; protein; 262 AA.  
XX  
XX AAE38640;  
AC  
XX  
DT 04-DEC-2003 (first entry)  
XX  
DE Human collagen type I cross-linked C-terminal peptide, PIIICP(alpha).  
XX  
KW Human; therapy; osteoporosis; left-ventricular hypertrophy; scleroderma;  
KW liver cirrhosis; congestive heart failure; multiple sclerosis; fibrosis;  
KW coal-miner's pneumoconiosis; muscle-wasting syndrome; diabetes mellitus;  
KW rheumatoid arthritis; Alzheimer's disease; muscular dystrophy; cancer;  
KW collagen type I cross-linked carboxy-terminal peptide.  
XX  
OS Homo sapiens.  
XX  
XX WO2003068919-A2.  
PN  
XX  
PD 21-AUG-2003.  
XX  
XX 12-FEB-2003; 2003WO-US004183.  
PF  
XX  
PR 12-FEB-2002; 2002US-0356008P.  
XX  
XX (REGC ) UNIV CALIFORNIA.  
PA  
XX  
XX Hellerstein MK;  
PI  
XX  
XX WPI; 2003-689661/65.  
DR  
XX  
XX Determining rate of biosynthesis or breakdown of inaccessible biological  
XX molecules, useful e.g. for diagnosis or monitoring treatment, by  
XX administering labeled precursor.  
XX  
XX Claim 24; Page 102-103; 105pp; English.  
XX  
XX The invention relates to a method of determining the rate of biosynthesis  
XX or breakdown of at least one inaccessible biological molecule in a  
XX subject. The method is useful for diagnosis or monitoring, and treatment  
XX of diseases associated with an altered rate of biosynthesis/breakdown of  
XX an isotopically labelled precursor molecule, specifically osteoporosis;  
XX left-ventricular hypertrophy; liver cirrhosis or fibrosis; congestive  
XX heart failure; scleroderma; coal-miner's pneumoconiosis; cardiac or lung  
XX fibrosis; Alzheimer's disease; multiple sclerosis; rheumatoid arthritis;  
XX diabetes mellitus; muscle-wasting syndromes; muscular dystrophy; athletic  
XX training and cancer. The method is also useful for screening candidate  
XX gene or protein targets, phenotypic/human validation studies on potential  
XX drugs, drug mechanism studies and determining the risk of developing the  
XX disease. The present sequence is human collagen type I cross-linked  
XX carboxy terminal peptide (ICTP). This sequence is used to illustrate the  
XX method of the invention  
XX  
XX Sequence 262 AA;  
SQ  
Query Match 100.0%; Score 42; DB 7; Length 262;  
Best Local Similarity 100.0%; Pred. No. 14;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 1 IGGEXAGG 8  
Db 4 IGGEXAGG 11  
|||||

RESULT 8  
AAE07375  
ID AAE07375 standard; protein; 270 AA.  
XX  
XX AAE07375;  
AC  
XX  
XX 16-JUL-1999 (first entry)  
DT  
XX

DE Procollagen-III-C-terminal propeptide.  
 XX Alcoholic cirrhosis; biliary cirrhosis; hepatitis; schistosomiasis;  
 KW cardiac fibrosis; Crohn's disease; diabetic nephropathy; collagen;  
 KW fibril; procollagen-III-C-terminal propeptide; immunoassay; diagnosis;  
 KW glucocorticosteroid; antibody.  
 XX Homo sapiens.  
 OS  
 XX  
 XX EP913692-A1.  
 PN  
 XX 06-MAY-1999.  
 PD  
 XX  
 XX 31-OCT-1997; 97EP-00119018.  
 PF  
 XX  
 XX 31-OCT-1997; 97EP-00119018.  
 PR  
 XX  
 XX (FARB ) BAYER AG.  
 PA  
 XX  
 XX Burchardt ER, Kroll W, Neumann R, Schroeder W;  
 PI  
 XX  
 XX WPI; 1999-256748/22.  
 DR  
 XX  
 XX N-PSDB; AAX57501.  
 PA  
 XX  
 XX New immunoassay for procollagen-III-C-terminal propeptide (PIIICP) useful  
 PT for diagnosing diseases such as alcoholic cirrhosis, hepatitis and  
 PT Duchenne's muscular dystrophy.  
 PT  
 XX  
 XX Disclosure; Fig 1; 28pp; English.  
 PS  
 XX Diseases such as alcoholic cirrhosis, biliary cirrhosis, hepatitis,  
 XX schistosomiasis, cardiac fibrosis, Crohn's disease, diabetic nephropathy  
 CC and fibroses caused by surgery result from an inappropriate production of  
 CC collagen fibrils. Collagen III is synthesised as a preproprotein which is  
 CC modified by post-translation modification. The mature collagen III  
 CC fibrils are generated by cleavage of a propeptide from the C-terminus  
 CC followed by cleavage of a fragment from the N-terminus. This sequence  
 CC corresponds to the procollagen III C-terminal propeptide (PIIICP). As the  
 CC fibrils are formed, the PIIICP propeptide is released into the  
 CC surrounding tissue fluids. The invention relates to an immunoassay to  
 CC determine PIIICP concentrations and thus diagnose the above diseases and  
 CC others which result from inappropriate collagen deposition. Different  
 CC fibrotic diseases may show different levels of PIIICP in the serum so the  
 CC immunoassay may also be used to determine which disease a particular  
 CC patient is suffering from. The new immunoassay may also be used to  
 CC measure the rate of collagen synthesis in patients being treated with  
 CC glucocorticosteroids. In addition the new antibodies may be used in  
 CC immunohistochemical staining of cryostat and paraffin sections to assess  
 CC collagen synthesis in tissue samples from patients suspected of having  
 CC fibrotic disease  
 XX  
 XX Sequence 270 AA;  
 SQ  
 Query Match 100.0%; Score 42; DB 2; Length 270;  
 Best Local Similarity 100.0%; Pred. No. 15;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 IGGEKAGG 8  
 Db 12 IGGEKAGG 19  
 |||||  
 RESULT 9  
 ABU70813  
 ID ABU70813 standard; protein; 293 AA.  
 XX  
 XX ABU70813;  
 AC  
 XX 10-JUN-2003 (first entry)  
 DT  
 XX Human adipocyte Selected Interacting domain, SID, #444.  
 DE  
 XX Human; prey; adipocyte; SID; selected interacting domain; anorectic;  
 KW Human; prey; adipocyte; SID; selected interacting domain; anorectic;

KW antidiabetic; protein-protein interaction; diabetes;  
 KW Yeast 2-hybrid assay; metabolic disorder; obesity.  
 XX  
 OS Homo sapiens.  
 XX  
 XX WO200286122-A2.  
 PN  
 XX 31-OCT-2002.  
 PD  
 XX  
 XX 14-MAR-2002; 2002WO-EP003768.  
 PF  
 XX  
 XX 14-MAR-2001; 2001US-0275734P.  
 PR  
 XX  
 XX (HYBR-) HYBRIGENICS.  
 PA  
 XX  
 XX Legrain P, Daviet L;  
 PI  
 XX  
 XX WPI; 2003-103412/09.  
 DR  
 XX  
 XX N-PSDB; ACA57357.  
 DR  
 XX  
 XX New complex between two interacting proteins in adipocyte cells, useful  
 PT for identifying selected interacting domains that modulate protein  
 PT interactions, or for preventing or treating metabolic disorders such as  
 PT obesity or diabetes.  
 PT  
 XX  
 XX Claim 6; Page 258-259; 382pp; English.  
 PS  
 XX The invention relates to a complex between two interacting proteins in  
 CC adipocyte cells, given in the specification. The proteins are identified  
 CC by selecting a bait protein from a known adipocyte marker and then  
 CC performing a yeast 2-hybrid selection to isolate prey proteins encoded by  
 CC members of an adipocyte cDNA library. The proteins are designated SID  
 CC (RTM) (selected interacting domains) proteins. Also included are a  
 CC polynucleotide encoding a polypeptide in the adipocyte cells, a  
 CC recombinant host cell expressing at least one of the interacting  
 CC polypeptides of the complex, selecting a modulating compound in adipocyte  
 CC cells, a SID (RTM) polypeptide comprising any of the 738 amino acid  
 CC sequences given in the specification (including its fragment or variant),  
 CC a SID (RTM) polynucleotide comprising any of the 738 nucleotide sequences  
 CC given in the specification (including its fragment or variant), a vector  
 CC comprising the SID (RTM) polynucleotide, a recombinant host cell  
 CC comprising the vector, a protein chip comprising the polypeptides and a  
 CC record comprising all or part of the data, listed in the specification.  
 CC The complex, polypeptides, polynucleotides and compounds are useful for  
 CC preventing or treating metabolic disorders such as obesity or diabetes.  
 CC The polynucleotides are useful as probes or primers. The complex is  
 CC particularly useful for identifying selected interacting domains (SID  
 CC (RTM)) for screening drugs that modulate the protein interaction, thus  
 CC exhibiting the therapeutic effect. The present sequence represents a SID  
 CC (prey) protein of the invention  
 XX  
 XX Sequence 293 AA;  
 SQ  
 Query Match 100.0%; Score 42; DB 6; Length 293;  
 Best Local Similarity 100.0%; Pred. No. 16;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 IGGEKAGG 8  
 Db 139 IGGEKAGG 146  
 |||||  
 RESULT 10  
 AAU23675  
 ID AAU23675 standard; protein; 309 AA.  
 XX  
 XX AAU23675;  
 AC  
 XX 18-DEC-2001 (first entry)  
 DT  
 XX Novel human enzyme polypeptide #761.  
 DE  
 XX Human; oxidoreductase enzyme; transferase; hydrolase; lyase; isomerase;  
 KW Human; oxidoreductase enzyme; transferase; hydrolase; lyase; isomerase;

KW ligase; hyperproliferative disorder; immunodeficiency disorder;  
KW autoimmune disorder; neurological disorder; metabolic disorder;  
KW inflammatory disorder; cardiovascular disorder; reproductive disorder;  
KW blood-related disorder; infectious disorder; cytostatic; anti arthritic;  
nephrotropic; anticoagulant.  
XX  
OS Homo sapiens.  
XX  
PN WC200155301-A2.  
XX  
PD 02-AUG-2001.  
XX  
XX  
PF 17-JAN-2001; 2001WO-US001239.  
XX  
PR 31-JAN-2000; 2000US-0179065P.  
PR 04-FEB-2000; 2000US-0180628P.  
PR 24-FEB-2000; 2000US-0184664P.  
PR 02-MAR-2000; 2000US-0186350P.  
PR 16-MAR-2000; 2000US-0189874P.  
PR 17-MAR-2000; 2000US-0190076P.  
PR 18-APR-2000; 2000US-0198123P.  
PR 19-MAY-2000; 2000US-0205515P.  
PR 07-JUN-2000; 2000US-0209467P.  
PR 28-JUN-2000; 2000US-0214886P.  
PR 30-JUN-2000; 2000US-0215135P.  
PR 07-JUL-2000; 2000US-0216647P.  
PR 07-JUL-2000; 2000US-0216880P.  
PR 11-JUL-2000; 2000US-0217487P.  
PR 14-JUL-2000; 2000US-0217496P.  
PR 14-JUL-2000; 2000US-0218290P.  
PR 26-JUL-2000; 2000US-0220963P.  
PR 26-JUL-2000; 2000US-0220964P.  
PR 14-AUG-2000; 2000US-0224518P.  
PR 14-AUG-2000; 2000US-0224519P.  
PR 14-AUG-2000; 2000US-0225213P.  
PR 14-AUG-2000; 2000US-0225214P.  
PR 14-AUG-2000; 2000US-0225266P.  
PR 14-AUG-2000; 2000US-0225287P.  
PR 14-AUG-2000; 2000US-0225758P.  
PR 18-AUG-2000; 2000US-0225759P.  
PR 22-AUG-2000; 2000US-0226681P.  
PR 22-AUG-2000; 2000US-0226686P.  
PR 23-AUG-2000; 2000US-0227182P.  
PR 30-AUG-2000; 2000US-0227009P.  
PR 01-SEP-2000; 2000US-0228924P.  
PR 01-SEP-2000; 2000US-0229343P.  
PR 01-SEP-2000; 2000US-0229344P.  
PR 05-SEP-2000; 2000US-0229345P.  
PR 05-SEP-2000; 2000US-0229509P.  
PR 06-SEP-2000; 2000US-0229513P.  
PR 06-SEP-2000; 2000US-0230437P.  
PR 06-SEP-2000; 2000US-0230438P.  
PR 08-SEP-2000; 2000US-0231242P.  
PR 08-SEP-2000; 2000US-0231243P.  
PR 08-SEP-2000; 2000US-0231244P.  
PR 08-SEP-2000; 2000US-0231413P.  
PR 08-SEP-2000; 2000US-0231414P.  
PR 08-SEP-2000; 2000US-0232080P.  
PR 08-SEP-2000; 2000US-0232081P.  
PR 12-SEP-2000; 2000US-0231968P.  
PR 14-SEP-2000; 2000US-0232397P.  
PR 14-SEP-2000; 2000US-0232398P.  
PR 14-SEP-2000; 2000US-0232399P.  
PR 14-SEP-2000; 2000US-0232400P.  
PR 14-SEP-2000; 2000US-0232401P.  
PR 14-SEP-2000; 2000US-0233063P.  
PR 14-SEP-2000; 2000US-0233064P.  
PR 14-SEP-2000; 2000US-0233065P.  
PR 21-SEP-2000; 2000US-0234223P.  
PR 21-SEP-2000; 2000US-0234274P.  
PR 25-SEP-2000; 2000US-0234997P.  
PR 26-SEP-2000; 2000US-0234988P.  
PR 27-SEP-2000; 2000US-0235834P.  
PR 27-SEP-2000; 2000US-0235836P.  
PR 29-SEP-2000; 2000US-0236367P.  
PR 29-SEP-2000; 2000US-0236368P.  
PR 29-SEP-2000; 2000US-0236369P.  
PR 29-SEP-2000; 2000US-0236370P.  
PR 02-OCT-2000; 2000US-0236802P.  
PR 02-OCT-2000; 2000US-0237037P.  
PR 02-OCT-2000; 2000US-0237038P.  
PR 02-OCT-2000; 2000US-0237039P.  
PR 02-OCT-2000; 2000US-0237040P.  
PR 13-OCT-2000; 2000US-0239935P.  
PR 13-OCT-2000; 2000US-0239937P.  
PR 20-OCT-2000; 2000US-0240960P.  
PR 20-OCT-2000; 2000US-0241221P.  
PR 20-OCT-2000; 2000US-0241785P.  
PR 20-OCT-2000; 2000US-0241786P.  
PR 20-OCT-2000; 2000US-0241787P.  
PR 20-OCT-2000; 2000US-0241808P.  
PR 20-OCT-2000; 2000US-0241809P.  
PR 20-OCT-2000; 2000US-0241826P.  
PR 08-NOV-2000; 2000US-0244617P.  
PR 08-NOV-2000; 2000US-0246474P.  
PR 08-NOV-2000; 2000US-0246475P.  
PR 08-NOV-2000; 2000US-0246476P.  
PR 08-NOV-2000; 2000US-0246477P.  
PR 08-NOV-2000; 2000US-0246478P.  
PR 08-NOV-2000; 2000US-0246523P.  
PR 08-NOV-2000; 2000US-0246524P.  
PR 08-NOV-2000; 2000US-0246525P.  
PR 08-NOV-2000; 2000US-0246526P.  
PR 08-NOV-2000; 2000US-0246527P.  
PR 08-NOV-2000; 2000US-0246528P.  
PR 08-NOV-2000; 2000US-0246532P.  
PR 08-NOV-2000; 2000US-0246609P.  
PR 08-NOV-2000; 2000US-0246610P.  
PR 08-NOV-2000; 2000US-0246611P.  
PR 08-NOV-2000; 2000US-0246613P.  
PR 17-NOV-2000; 2000US-0249207P.  
PR 17-NOV-2000; 2000US-0249208P.  
PR 17-NOV-2000; 2000US-0249209P.  
PR 17-NOV-2000; 2000US-0249210P.  
PR 17-NOV-2000; 2000US-0249211P.  
PR 17-NOV-2000; 2000US-0249212P.  
PR 17-NOV-2000; 2000US-0249213P.  
PR 17-NOV-2000; 2000US-0249214P.  
PR 17-NOV-2000; 2000US-0249215P.  
PR 17-NOV-2000; 2000US-0249216P.  
PR 17-NOV-2000; 2000US-0249217P.  
PR 17-NOV-2000; 2000US-0249218P.  
PR 17-NOV-2000; 2000US-0249244P.  
PR 17-NOV-2000; 2000US-0249245P.  
PR 17-NOV-2000; 2000US-0249264P.  
PR 17-NOV-2000; 2000US-0249265P.  
PR 17-NOV-2000; 2000US-0249297P.  
PR 17-NOV-2000; 2000US-0249299P.  
PR 17-NOV-2000; 2000US-0249300P.  
PR 01-DEC-2000; 2000US-0250160P.  
PR 01-DEC-2000; 2000US-0250391P.  
PR 05-DEC-2000; 2000US-0251030P.  
PR 05-DEC-2000; 2000US-0251988P.  
PR 06-DEC-2000; 2000US-0256719P.  
PR 05-DEC-2000; 2000US-0251479P.  
PR 08-DEC-2000; 2000US-0251856P.  
PR 08-DEC-2000; 2000US-0251858P.  
PR 08-DEC-2000; 2000US-0251869P.

XX	08-DEC-2000; 2000US-0251989P.
PD	08-DEC-2000; 2000US-0251990P.
XX	11-DEC-2000; 2000US-0254097F.
XX	05-JAN-2001; 2001US-0259678P.
XX	(HUMA-) HUMAN GENOME SCI INC.
XX	Rosen CA, Barash SC, Ruben SM;
PI	WPI; 2001-465566/50.
DR	N-PSDB; AAS41545.
XX	Novel polypeptides and polynucleotides useful for diagnosing, preventing,
PT	treating neural, immune system, muscular, reproductive, pulmonary,
PT	cardiovascular, renal, proliferative disorders and cancerous diseases.
XX	Claim 11; SEQ ID NO 1671; 1180pp; English.
XX	The present invention relates to the isolation of novel human enzyme
CC	polypeptides, and the cDNA (AAS40785-AAS41684) and genomic sequences
CC	encoding them. The enzyme polypeptides of the invention may comprise the
CC	functional classes of oxidoreductases, transferases, hydrolases, lyases,
CC	isomerases or ligases. The sequences of the invention are useful in the
CC	diagnosis, treatment, prevention and/or prognosis of a wide range of
CC	disorders, including hyperproliferative disorders (e.g. cancer),
CC	immunodeficiency disorders (e.g. AIDS) autoimmune disorders (e.g.
CC	arthritis), neurological disorders (e.g. Alzheimer's disease), metabolic
CC	disorders (e.g. phenylketonuria), inflammatory disorders (e.g. asthma),
CC	cardiovascular disorders (e.g. atherosclerosis), blood-related disorders
CC	(e.g. haemophilia), reproductive disorders (e.g. infertility) and
CC	infectious disorders (e.g. influenza). The polynucleotides of the
CC	invention can also be used in gene therapy. AAU22915-AAU23814 represent
CC	the novel human enzyme polypeptides of the invention. Note: The sequence
CC	data for this patent did not form part of the printed specification, but
CC	was obtained in electronic format directly from WIPO at
CC	ftp.wipo.int/pub/published_pct_sequences
XX	Sequence 309 AA;
SQ	Query Match 100.0%; Score 42; DB 4; Length 309;
	Best Local Similarity 100.0%; Pred. No. 17;
	Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY	1 IGGERKAGG 8
Dh	
Db	51 IGGERKAGG 58
RESULT 11	
ID	ABG60248 standard; protein; 309 AA.
AC	ABG60248;
XX	13-AUG-2002 (first entry)
DT	Human ovarian antigen #10.
DE	Human; ovarian antigen; ovary disorder; breast disorder;
XX	neoplastic disorder; cancer; infectious disease; inflammatory disease;
KW	reproductive system disorder; autoimmune disorder; Alzheimer's disease;
KW	blood-related disorder; hyperproliferative disorder; hair loss;
KW	urinary system disorder; cardiovascular disorder; arrhythmia;
KW	respiratory disorder; musculoskeletal system disorder;
KW	neural activity disorder; neurological disorder; endocrine disorder;
KW	gastrointestinal disorder; liver disorder; pancreatic disorder;
KW	gall bladder disorder; large intestine disorder; developmental disorder;
KW	inherited disorder; wound healing; skin aging; food additive;
XX	preservative.
OS	Homo sapiens.
XX	WC0200155329-A2.
XX	
XX	02-AUG-2001.
XX	17-JAN-2001; 2001WO-US001360.
XX	31-JAN-2000; 2000US-0179065P.
PR	04-FEB-2000; 2000US-0180628P.
PR	07-JUN-2000; 2000US-0209467P.
PR	14-SEP-2000; 2000US-0232398P.
PR	17-NOV-2000; 2000US-0249300P.
PR	01-DEC-2000; 2000US-0250160P.
PR	08-DEC-2000; 2000US-0251868P.
PR	08-DEC-2000; 2000US-0251990P.
XX	(HUMA-) HUMAN GENOME SCI INC.
XX	Rosen CA, Barash SC, Ruben SM;
PI	WPI; 2001-476195/51.
DR	N-PSDB; ABK72051.
XX	Novel isolated human ovarian related polypeptide useful for
PT	diagnosis/treatment of disorders of ovary and breast such as neoplastic
PT	disorders, infectious diseases, inflammatory diseases, and reproductive
PT	disorders.
XX	Claim 11; SEQ ID NO 78; 524pp; English.
PS	The invention relates to isolated ovarian related polypeptide (ovarian
XX	antigen) comprising a sequence at least 90% identical to a sequence
CC	selected from a polypeptide fragment, domain, epitope or full length
CC	protein of a sequence (S1) appearing as ABG60239-ABG60296 having
CC	biological activity, or a variant, allelic variant or species homologue
CC	of S1. Also included are the cDNA clones encoding the proteins of S1. S1,
CC	an anti-S1 antibody and the cDNA are useful for diagnosing, preventing,
CC	treating or ameliorating a medical condition in mammalian subject
CC	especially diseases and/or disorders of the ovary and/or breast such as
CC	neoplastic disorders (such as ovarian Krukenberg tumour and cancer),
CC	infectious diseases (e.g., mastitis, oophoritis), inflammatory diseases
CC	(e.g., abscesses), reproductive system disorders (Paget's disease),
CC	autoimmune disorders (systemic lupus erythematosus, rheumatoid
CC	arthritis), blood-related disorders (sickle cell anaemia),
CC	hyperproliferative disorders, urinary system disorders
CC	(glomerulonephritis), cardiovascular disorders (arrhythmias), respiratory
CC	disorders, musculoskeletal system disorders, neural activity and
CC	neurological disorders (Alzheimer's disease and Parkinson's disease),
CC	endocrine disorders (Addison's disease), gastrointestinal disorders
CC	(inflammatory disorders), liver disorders (biliary liver cirrhosis),
CC	pancreatic and gall bladder disorders, disorders of the large intestine,
CC	developmental and inherited disorders, diseases at the cellular level,
CC	and wound healing and epithelial cell proliferation. They are also useful
CC	to prevent skin aging, for preventing hair loss, to maintain organs
CC	before transplantation or for supporting cell culture of primary tissues,
CC	to modulate mammalian characteristics such as body height, to modulate
CC	mammalian metabolism, to change a mammal's mental or physical state, and
CC	as food additive or preservative. The present sequence represents an
CC	ovarian antigen, S1 protein of the invention. Note: The sequence data for
CC	this patent did not form part of the printed specification, but was
CC	obtained in electronic format directly from WIPO at
CC	ftp.wipo.int/pub/published_pct_sequences
XX	Sequence 309 AA;
SQ	Query Match 100.0%; Score 42; DB 4; Length 309;
	Best Local Similarity 100.0%; Pred. No. 17;
	Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY	1 IGGERKAGG 8
Dh	
Db	51 IGGERKAGG 58

XX	08-DEC-2000; 2000US-0251989P.
PD	08-DEC-2000; 2000US-0251990P.
XX	11-DEC-2000; 2000US-0254097F.
XX	05-JAN-2001; 2001US-0259678P.
XX	(HUMA-) HUMAN GENOME SCI INC.
XX	Rosen CA, Barash SC, Ruben SM;
PI	WPI; 2001-465566/50.
DR	N-PSDB; AAS41545.
XX	Novel polypeptides and polynucleotides useful for diagnosing, preventing,
PT	treating neural, immune system, muscular, reproductive, pulmonary,
PT	cardiovascular, renal, proliferative disorders and cancerous diseases.
XX	Claim 11; SEQ ID NO 1671; 1180pp; English.
XX	The present invention relates to the isolation of novel human enzyme
CC	polypeptides, and the cDNA (AAS40785-AAS41684) and genomic sequences
CC	encoding them. The enzyme polypeptides of the invention may comprise the
CC	functional classes of oxidoreductases, transferases, hydrolases, lyases,
CC	isomerases or ligases. The sequences of the invention are useful in the
CC	diagnosis, treatment, prevention and/or prognosis of a wide range of
CC	disorders, including hyperproliferative disorders (e.g. cancer),
CC	immunodeficiency disorders (e.g. AIDS) autoimmune disorders (e.g.
CC	arthritis), neurological disorders (e.g. Alzheimer's disease), metabolic
CC	disorders (e.g. phenylketonuria), inflammatory disorders (e.g. asthma),
CC	cardiovascular disorders (e.g. atherosclerosis), blood-related disorders
CC	(e.g. haemophilia), reproductive disorders (e.g. infertility) and
CC	infectious disorders (e.g. influenza). The polynucleotides of the
CC	invention can also be used in gene therapy. AAU22915-AAU23814 represent
CC	the novel human enzyme polypeptides of the invention. Note: The sequence
CC	data for this patent did not form part of the printed specification, but
CC	was obtained in electronic format directly from WIPO at
CC	ftp.wipo.int/pub/published_pct_sequences
XX	Sequence 309 AA;
SQ	Query Match 100.0%; Score 42; DB 4; Length 309;
	Best Local Similarity 100.0%; Pred. No. 17;
	Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY	1 IGGERKAGG 8
Dh	
Db	51 IGGERKAGG 58
RESULT 11	
ID	ABG60248 standard; protein; 309 AA.
AC	ABG60248;
XX	13-AUG-2002 (first entry)
DT	Human ovarian antigen #10.
DE	Human; ovarian antigen; ovary disorder; breast disorder;
XX	neoplastic disorder; cancer; infectious disease; inflammatory disease;
KW	reproductive system disorder; autoimmune disorder; Alzheimer's disease;
KW	blood-related disorder; hyperproliferative disorder; hair loss;
KW	urinary system disorder; cardiovascular disorder; arrhythmia;
KW	respiratory disorder; musculoskeletal system disorder;
KW	neural activity disorder; neurological disorder; endocrine disorder;
KW	gastrointestinal disorder; liver disorder; pancreatic disorder;
KW	gall bladder disorder; large intestine disorder; developmental disorder;
KW	inherited disorder; wound healing; skin aging; food additive;
XX	preservative.
OS	Homo sapiens.
XX	WC0200155329-A2.
XX	
XX	02-AUG-2001.
XX	17-JAN-2001; 2001WO-US001360.
XX	31-JAN-2000; 2000US-0179065P.
PR	04-FEB-2000; 2000US-0180628P.
PR	07-JUN-2000; 2000US-0209467P.
PR	14-SEP-2000; 2000US-0232398P.
PR	17-NOV-2000; 2000US-0249300P.
PR	01-DEC-2000; 2000US-0250160P.
PR	08-DEC-2000; 2000US-0251868P.
PR	08-DEC-2000; 2000US-0251990P.
PA	(HUMA-) HUMAN GENOME SCI INC.
XX	Rosen CA, Barash SC, Ruben SM;
PI	WPI; 2001-476195/51.
DR	N-PSDB; ABK72051.
XX	Novel isolated human ovarian related polypeptide useful for
PT	diagnosis/treatment of disorders of ovary and breast such as neoplastic
PT	disorders, infectious diseases, inflammatory diseases, and reproductive
PT	disorders.
XX	Claim 11; SEQ ID NO 78; 524pp; English.
PS	The invention relates to isolated ovarian related polypeptide (ovarian
XX	antigen) comprising a sequence at least 90% identical to a sequence
CC	selected from a polypeptide fragment, domain, epitope or full length
CC	protein of a sequence (S1) appearing as ABG60239-ABG60296 having
CC	biological activity, or a variant, allelic variant or species homologue
CC	of S1. Also included are the cDNA clones encoding the proteins of S1. S1,
CC	an anti-S1 antibody and the cDNA are useful for diagnosing, preventing,
CC	treating or ameliorating a medical condition in mammalian subject
CC	especially diseases and/or disorders of the ovary and/or breast such as
CC	neoplastic disorders (such as ovarian Krukenberg tumour and cancer),
CC	infectious diseases (e.g., mastitis, oophoritis), inflammatory diseases
CC	(e.g., abscesses), reproductive system disorders (Paget's disease),
CC	autoimmune disorders (systemic lupus erythematosus, rheumatoid
CC	arthritis), blood-related disorders (sickle cell anaemia),
CC	hyperproliferative disorders, urinary system disorders
CC	(glomerulonephritis), cardiovascular disorders (arrhythmias), respiratory
CC	disorders, musculoskeletal system disorders, neural activity and
CC	neurological disorders (Alzheimer's disease and Parkinson's disease),
CC	endocrine disorders (Addison's disease), gastrointestinal disorders
CC	(inflammatory disorders), liver disorders (biliary liver cirrhosis),
CC	pancreatic and gall bladder disorders, disorders of the large intestine,
CC	developmental and inherited disorders, diseases at the cellular level,
CC	and wound healing and epithelial cell proliferation. They are also useful
CC	to prevent skin aging, for preventing hair loss, to maintain organs
CC	before transplantation or for supporting cell culture of primary tissues,
CC	to modulate mammalian characteristics such as body height, to modulate
CC	mammalian metabolism, to change a mammal's mental or physical state, and
CC	as food additive or preservative. The present sequence represents an
CC	ovarian antigen, S1 protein of the invention. Note: The sequence data for
CC	this patent did not form part of the printed specification, but was
CC	obtained in electronic format directly from WIPO at
CC	ftp.wipo.int/pub/published_pct_sequences
XX	Sequence 309 AA;
SQ	Query Match 100.0%; Score 42; DB 4; Length 309;
	Best Local Similarity 100.0%; Pred. No. 17;
	Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY	1 IGGERKAGG 8
Dh	
Db	51 IGGERKAGG 58

RESULT 12  
 ABG61719  
 ID ABG61719 standard; protein; 309 AA.  
 XX  
 AC ABG61719;  
 XX  
 DT 26-AUG-2002 (first entry)  
 XX  
 DE Novel ovarian related polypeptide #10.  
 XX  
 KW Ovarian related polypeptide; neoplastic disorder; tumour; ovarian cancer;  
 KW hyperproliferative disorder; adult acute lymphocytic leukaemia;  
 KW breast cancer; reproductive system disorder; tuberculosis; arthritis;  
 KW immune system disorder; Chediak-Higashi's syndrome; neonatal neutropenia;  
 KW autoimmune disorder; Hashimoto's thyroiditis; inflammatory disorder;  
 KW septic shock; multiple sclerosis; central nervous system disorder;  
 KW neurological disorder; allergy; Parkinson's disease; Alzheimer's disease;  
 KW cardiovascular disorder; atherosclerosis; blood related disorder;  
 KW respiratory disorder; urinary system disorder; musculoskeletal disorder;  
 KW osteoporosis; wound healing; endocrine disorder; infectious disease;  
 KW gastrointestinal disorder; transplantation; food additive; preservative.  
 XX  
 OS Homo sapiens.  
 XX  
 XX US2002045230-A1.  
 XX  
 PD 18-APR-2002.  
 XX  
 XX 20-JUL-2001; 2001US-00908711.  
 XX  
 PR 31-JAN-2000; 2000US-0179065P.  
 PR 04-FEB-2000; 2000US-0180628P.  
 PR 24-FEB-2000; 2000US-0184664P.  
 PR 02-MAR-2000; 2000US-0186350P.  
 PR 16-MAR-2000; 2000US-0189874P.  
 PR 17-MAR-2000; 2000US-0190076P.  
 PR 18-APR-2000; 2000US-0198123P.  
 PR 19-MAY-2000; 2000US-0205515P.  
 PR 07-JUN-2000; 2000US-0209467P.  
 PR 28-JUN-2000; 2000US-0214886P.  
 PR 30-JUN-2000; 2000US-0215135P.  
 PR 07-JUL-2000; 2000US-0216647P.  
 PR 07-JUL-2000; 2000US-0216880P.  
 PR 11-JUL-2000; 2000US-0217487P.  
 PR 11-JUL-2000; 2000US-0217496P.  
 PR 14-JUL-2000; 2000US-0218290P.  
 PR 26-JUL-2000; 2000US-0220963P.  
 PR 26-JUL-2000; 2000US-0220964P.  
 PR 14-AUG-2000; 2000US-0224519P.  
 PR 14-AUG-2000; 2000US-0225213P.  
 PR 14-AUG-2000; 2000US-0225214P.  
 PR 14-AUG-2000; 2000US-0225266P.  
 PR 14-AUG-2000; 2000US-0225267P.  
 PR 14-AUG-2000; 2000US-0225268P.  
 PR 14-AUG-2000; 2000US-0225270P.  
 PR 14-AUG-2000; 2000US-0225447P.  
 PR 14-AUG-2000; 2000US-0225757P.  
 PR 14-AUG-2000; 2000US-0225758P.  
 PR 14-AUG-2000; 2000US-0225759P.  
 PR 18-AUG-2000; 2000US-0226279P.  
 PR 22-AUG-2000; 2000US-0226681P.  
 PR 22-AUG-2000; 2000US-0226868P.  
 PR 22-AUG-2000; 2000US-0227182P.  
 PR 23-AUG-2000; 2000US-0227009P.  
 PR 30-AUG-2000; 2000US-0228924P.  
 PR 01-SEP-2000; 2000US-0229287P.  
 PR 01-SEP-2000; 2000US-0229343P.  
 PR 01-SEP-2000; 2000US-0229344P.  
 PR 01-SEP-2000; 2000US-0229345P.  
 PR 05-SEP-2000; 2000US-0229509P.  
 PR 05-SEP-2000; 2000US-0229513P.  
 PR 06-SEP-2000; 2000US-0230437P.  
 PR 06-SEP-2000; 2000US-0230438P.  
 PR 08-SEP-2000; 2000US-0231242P.  
 PR 08-SEP-2000; 2000US-0231243P.  
 PR 08-SEP-2000; 2000US-0231244P.  
 PR 08-SEP-2000; 2000US-0231413P.  
 PR 08-SEP-2000; 2000US-0231414P.  
 PR 08-SEP-2000; 2000US-0232080P.  
 PR 08-SEP-2000; 2000US-0232081P.  
 PR 12-SEP-2000; 2000US-0231968P.  
 PR 14-SEP-2000; 2000US-0232397P.  
 PR 14-SEP-2000; 2000US-0232398P.  
 PR 14-SEP-2000; 2000US-0232399P.  
 PR 14-SEP-2000; 2000US-0232400P.  
 PR 14-SEP-2000; 2000US-0232401P.  
 PR 14-SEP-2000; 2000US-0233063P.  
 PR 14-SEP-2000; 2000US-0233064P.  
 PR 14-SEP-2000; 2000US-0233065P.  
 PR 21-SEP-2000; 2000US-0234223P.  
 PR 21-SEP-2000; 2000US-0234274P.  
 PR 21-SEP-2000; 2000US-0234997P.  
 PR 25-SEP-2000; 2000US-0234998P.  
 PR 26-SEP-2000; 2000US-0235484P.  
 PR 27-SEP-2000; 2000US-0235834P.  
 PR 27-SEP-2000; 2000US-0235836P.  
 PR 29-SEP-2000; 2000US-0236327P.  
 PR 29-SEP-2000; 2000US-0236367P.  
 PR 29-SEP-2000; 2000US-0236368P.  
 PR 29-SEP-2000; 2000US-0236369P.  
 PR 29-SEP-2000; 2000US-0236370P.  
 PR 02-OCT-2000; 2000US-0236802P.  
 PR 02-OCT-2000; 2000US-0237037P.  
 PR 02-OCT-2000; 2000US-0237038P.  
 PR 02-OCT-2000; 2000US-0237039P.  
 PR 02-OCT-2000; 2000US-0237040P.  
 PR 13-OCT-2000; 2000US-0239935P.  
 PR 13-OCT-2000; 2000US-0239937P.  
 PR 20-OCT-2000; 2000US-0240950P.  
 PR 20-OCT-2000; 2000US-0241221P.  
 PR 20-OCT-2000; 2000US-0241785P.  
 PR 20-OCT-2000; 2000US-0241786P.  
 PR 20-OCT-2000; 2000US-0241787P.  
 PR 20-OCT-2000; 2000US-0241808P.  
 PR 20-OCT-2000; 2000US-0241809P.  
 PR 20-OCT-2000; 2000US-0241826P.  
 PR 01-NOV-2000; 2000US-0244617P.  
 PR 08-NOV-2000; 2000US-0246474P.  
 PR 08-NOV-2000; 2000US-0246475P.  
 PR 08-NOV-2000; 2000US-0246476P.  
 PR 08-NOV-2000; 2000US-0246477P.  
 PR 08-NOV-2000; 2000US-0246478P.  
 PR 08-NOV-2000; 2000US-0246523P.  
 PR 08-NOV-2000; 2000US-0246524P.  
 PR 08-NOV-2000; 2000US-0246525P.  
 PR 08-NOV-2000; 2000US-0246526P.  
 PR 08-NOV-2000; 2000US-0246537P.  
 PR 08-NOV-2000; 2000US-0246528P.  
 PR 08-NOV-2000; 2000US-0246532P.  
 PR 08-NOV-2000; 2000US-0246609P.  
 PR 08-NOV-2000; 2000US-0246610P.  
 PR 08-NOV-2000; 2000US-0246611P.  
 PR 08-NOV-2000; 2000US-0246613P.  
 PR 17-NOV-2000; 2000US-0249207P.  
 PR 17-NOV-2000; 2000US-0249208P.  
 PR 17-NOV-2000; 2000US-0249209P.  
 PR 17-NOV-2000; 2000US-0249210P.  
 PR 17-NOV-2000; 2000US-0249211P.  
 PR 17-NOV-2000; 2000US-0249212P.  
 PR 17-NOV-2000; 2000US-0249213P.  
 PR 17-NOV-2000; 2000US-0249214P.  
 PR 17-NOV-2000; 2000US-0249215P.  
 PR 17-NOV-2000; 2000US-0249216P.





XX  
DT 15-DEC-1997 (first entry)  
DE  
DE Pro-alpha(III):(I) CP chimeric protein.  
XX  
XX C-propeptide; recognition sequence; procollagen; monomer chain; therapy;  
KW trimerised pro-alpha chain; fibril; procollagen suicide; wound healing;  
KW fibrotic disease; human; chimeric protein.  
XX  
OS Homo sapiens.  
XX  
XX WO9708311-A1.  
PN  
XX  
XX 06-MAR-1997.  
PD  
XX  
XX 30-AUG-1996; 96WO-GB002122.  
XX  
XX 31-AUG-1995; 95GB-00017773.  
PR  
XX 23-MAR-1996; 96GB-00006152.  
PR  
XX 14-JUN-1996; 96GB-00012476.  
PR  
XX (UYMA-) UNIV VICTORIA MANCHESTER.  
PA  
XX  
XX Bulleid N, Kadler K;  
PI  
XX WPI; 1997-179268/16.  
XX  
XX Novel pro:collagen mol. - comprising pro:collagen C-pro:peptide attached  
PT to an alien collagen alpha-chain or non-collagen material, useful e.g.  
PT for wound healing.  
XX  
XX Example 1; Page 32-35; 69pp; English.  
FS  
XX This sequence represents a chimeric procollagen molecule of the  
CC invention. This sequence has the procollagen C-propeptide from the pro  
CC -alpha2(I) chain attached to the pro-alpha(III) chain sequence. The C-  
CC propeptide is implicated in the assembly of the monomer chains into  
CC trimerised pro-alpha chains prior to cleavage of the N- and C-propeptides  
CC and formation of collagen in fibril-forming pro-alpha chains. The C-  
CC propeptides determine the type-specific assembly of the moieties to which  
CC they are attached. The molecule of the invention comprises a first moiety  
CC having procollagen C-propeptide activity attached to a second moiety,  
CC which is an alien collagen alpha-chain or a non-collagen material. The  
CC novel collagen molecule can be used for treatment or diagnosis in humans  
CC or animals, especially for the treatment of procollagen suicide, as an  
CC adhesive or implant, to promote (chronic) wound healing or fibrotic  
CC diseases with reduced scarring or for use in photography, brewing,  
CC foodstuffs or textiles. The novel collagen molecules, especially when  
CC containing substitutions in the recognition site, may have significantly  
CC altered properties and characteristics, such as different binding  
CC kinetics or alpha-chain selection properties  
XX  
SQ Sequence 623 AA;  
Query Match 100.0%; Score 42; DB 2; Length 623;  
Best Local Similarity 100.0%; Pred. No. 36;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 IGGEKAGG 8  
Db 368 IGGEKAGG 375  
RESULT 15  
AAW12842  
ID AAW12842 standard; protein; 626 AA.  
XX  
XX AAW12842;  
AC  
XX 15-DEC-1997 (first entry)  
DT  
XX Truncated pro-alpha(III) chain.  
DE  
XX

KW  
KW C-propeptide; recognition sequence; procollagen; monomer chain; therapy;  
KW trimerised pro-alpha chain; fibril; procollagen suicide; wound healing;  
XX fibrotic disease; human.  
OS Homo sapiens.  
XX  
XX WO9708311-A1.  
PN  
XX  
XX 06-MAR-1997.  
PD  
XX  
XX 30-AUG-1996; 96WO-GB002122.  
XX  
XX 31-AUG-1995; 95GB-00017773.  
PR  
XX 23-MAR-1996; 96GB-00006152.  
PR  
XX 14-JUN-1996; 96GB-00012476.  
PR  
XX (UYMA-) UNIV VICTORIA MANCHESTER.  
PA  
XX  
XX Bulleid N, Kadler K;  
PI  
XX WPI; 1997-179268/16.  
XX  
XX N-PSDB; AAT59892.  
DR  
XX Novel pro:collagen mol. - comprising pro:collagen C-pro:peptide attached  
PT to an alien collagen alpha-chain or non-collagen material, useful e.g.  
PT for wound healing.  
XX  
XX Example 1; Page 28-31; 69pp; English.  
FS  
XX This sequence represents a truncated procollagen pro-alpha(III) chain  
CC that can be used in the procollagen molecules of the invention. The C-  
CC propeptide is implicated in the assembly of the monomer chains into  
CC trimerised pro-alpha chains prior to cleavage of the N- and C-propeptides  
CC and formation of collagen in fibril-forming pro-alpha chains. The C-  
CC propeptides determine the type-specific assembly of the moieties to which  
CC they are attached. The molecule of the invention comprises a first moiety  
CC having procollagen C-propeptide activity attached to a second moiety,  
CC which is an alien collagen alpha-chain or a non-collagen material. The  
CC novel collagen molecule can be used for treatment or diagnosis in humans  
CC or animals, especially for the treatment of procollagen suicide, as an  
CC adhesive or implant, to promote (chronic) wound healing or fibrotic  
CC diseases with reduced scarring or for use in photography, brewing,  
CC foodstuffs or textiles. The novel collagen molecules, especially when  
CC containing substitutions in the recognition site, may have significantly  
CC altered properties and characteristics, such as different binding  
CC kinetics or alpha-chain selection properties  
XX  
SQ Sequence 626 AA;  
Query Match 100.0%; Score 42; DB 2; Length 626;  
Best Local Similarity 100.0%; Pred. No. 36;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 IGGEKAGG 8  
Db 368 IGGEKAGG 375  
Search completed: September 18, 2004, 04:25:31  
Job time : 43.8235 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 18, 2004, 04:20:54 ; Search time 13.1765 Seconds  
(without alignments)  
31.344 Million cell updates/sec

Title: US-10-615-959-40

Perfect score: 42

Sequence: 1 IGGEKAGG 8

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents AA:\*  
1: /cgn2\_6/ptodata/2/iaa/5A-COMB.pep:\*  
2: /cgn2\_6/ptodata/2/iaa/5B-COMB.pep:\*  
3: /cgn2\_6/ptodata/2/iaa/6A-COMB.pep:\*  
4: /cgn2\_6/ptodata/2/iaa/6B-COMB.pep:\*  
5: /cgn2\_6/ptodata/2/iaa/PCTUS-COMB.pep:\*  
6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	42	100.0	8	4	US-10-009-999A-40
2	42	100.0	9	4	US-10-009-999A-41
3	42	100.0	11	4	US-10-009-999A-42
4	42	100.0	12	4	US-10-009-999A-43
5	42	100.0	15	4	US-08-278-774-20
6	42	100.0	623	3	US-09-023-348-3
7	42	100.0	626	3	US-09-029-348-2
8	38	90.5	674	4	US-09-199-637A-51
9	38	90.5	674	4	US-09-252-991A-26476
10	36	85.7	11	6	5473052-20
11	36	85.7	1078	3	US-08-963-825-21
12	36	85.7	1078	4	US-09-500-811-21
13	36	85.7	1078	4	US-09-570-573-21
14	36	85.7	1078	4	US-09-548-608-21
15	35	83.3	561	4	US-09-252-991A-28074
16	34	81.0	529	4	US-09-252-991A-27659
17	34	81.0	540	4	US-09-252-991A-23300
18	34	81.0	822	4	US-09-252-991A-22479
19	33	78.6	221	4	US-09-252-991A-28805
20	33	78.6	254	4	US-09-252-991A-22621
21	33	78.6	311	2	US-08-602-359A-41
22	32	76.2	15	1	US-08-383-753-90
23	32	76.2	15	2	US-08-586-772-90
24	32	76.2	15	2	US-08-959-512-90
25	32	76.2	15	3	US-09-512-983-90
26	32	76.2	20	1	US-08-478-312-53
27	32	76.2	20	1	US-08-485-302-53

28	32	76.2	20	1	US-08-476-169-58	Sequence 58, Appl
29	32	76.2	20	1	US-08-484-083-58	Sequence 58, Appl
30	32	76.2	20	1	US-08-548-540-91	Sequence 91, Appl
31	32	76.2	20	5	PCT-US96-09809-91	Sequence 91, Appl
32	32	76.2	21	1	US-07-963-321-91	Sequence 91, Appl
33	32	76.2	21	1	US-08-290-641-91	Sequence 91, Appl
34	32	76.2	21	2	US-08-764-640-236	Sequence 236, App
35	32	76.2	21	3	US-08-973-225-224	Sequence 224, App
36	32	76.2	21	3	US-09-244-298A-236	Sequence 236, App
37	32	76.2	21	3	US-09-516-704-236	Sequence 236, App
38	32	76.2	21	4	US-09-549-090-224	Sequence 224, App
39	32	76.2	21	4	US-09-832-230A-236	Sequence 236, App
40	32	76.2	28	1	US-08-708-893-1	Sequence 1, Appl
41	32	76.2	28	3	US-09-088-154-1	Sequence 1, Appl
42	32	76.2	131	4	US-09-489-039A-10730	Sequence 10730, A
43	32	76.2	186	4	US-09-252-991A-27476	Sequence 27476, A
44	32	76.2	214	4	US-09-252-991A-31304	Sequence 31304, A
45	32	76.2	239	4	US-09-252-991A-30746	Sequence 30746, A

#### ALIGNMENTS

RESULT 1  
US-10-009-999A-40  
; Sequence 40, Application US/100099999A  
; Patent No. 6602980  
; GENERAL INFORMATION:  
; APPLICANT: Eyre, David R.  
; TITLE OF INVENTION: SYNTHETIC PEPTIDES FOR COLLAGEN RESORPTION ASSAYS  
; FILE REFERENCE: WROG-1-18220  
; CURRENT APPLICATION NUMBER: US/10/009,999A  
; CURRENT FILING DATE: 2003-01-22  
; PRIOR APPLICATION NUMBER: PCT/US99/29357  
; PRIOR FILING DATE: 1999-12-10  
; PRIOR APPLICATION NUMBER: US 09/335,098  
; PRIOR FILING DATE: 1999-06-17  
; PRIOR APPLICATION NUMBER: US 60/141,574  
; PRIOR FILING DATE: 1999-06-29  
; PRIOR APPLICATION NUMBER: US 60/142,274  
; PRIOR FILING DATE: 1999-07-02  
; PRIOR APPLICATION NUMBER: US 60/142,675  
; PRIOR FILING DATE: 1999-07-07  
; PRIOR APPLICATION NUMBER: US 09/385,740  
; PRIOR FILING DATE: 1999-08-30  
; NUMBER OF SEQ ID NOS: 45  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 40  
; LENGTH: 8  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: MISC FEATURE  
; OTHER INFORMATION: Synthetic peptide corresponding to carboxy-terminal  
; OTHER INFORMATION: telopeptide sequence of human type III collagen  
US-10-009-999A-40

Query Match 100.0%; Score 42; DB 4; Length 8;  
Best Local Similarity 100.0%; Pred. No. 3e+05;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QV 1 IGGEKAGG 8  
| | | | | | | |  
DB 1 IGGEKAGG 8

RESULT 2  
US-10-009-999A-41  
; Sequence 41, Application US/100099999A  
; Patent No. 6602980  
; GENERAL INFORMATION:  
; APPLICANT: Eyre, David R.  
; TITLE OF INVENTION: SYNTHETIC PEPTIDES FOR COLLAGEN RESORPTION ASSAYS

FILE REFERENCE: WROS-1-18220  
; CURRENT APPLICATION NUMBER: US/10/009,999A  
; CURRENT FILING DATE: 2003-01-22  
; PRIOR APPLICATION NUMBER: PCT/US99/29357  
; PRIOR FILING DATE: 1999-12-10  
; PRIOR APPLICATION NUMBER: US 09/335,098  
; PRIOR FILING DATE: 1999-06-17  
; PRIOR APPLICATION NUMBER: US 60/141,574  
; PRIOR FILING DATE: 1999-06-29  
; PRIOR APPLICATION NUMBER: US 60/142,274  
; PRIOR FILING DATE: 1999-07-02  
; PRIOR APPLICATION NUMBER: US 60/142,675  
; PRIOR FILING DATE: 1999-07-07  
; PRIOR APPLICATION NUMBER: US 09/385,740  
; PRIOR FILING DATE: 1999-08-30  
; NUMBER OF SEQ ID NOS: 45  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 41  
; LENGTH: 9  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: MISC FEATURE  
; OTHER INFORMATION: synthetic peptide corresponding to carboxy-terminal  
; OTHER INFORMATION: telopeptide sequence of human type III collagen  
US-10-009-999A-41

Query Match 100.0%; Score 42; DB 4; Length 9;  
Best Local Similarity 100.0%; Pred. No. 3e+05;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 IGGEKAGG 8  
| | | | |  
Db 1 IGGEKAGG 8

RESULT 3  
US-10-009-999A-42  
; Sequence 42, Application US/10009999A  
; Patent No. 6602980  
; GENERAL INFORMATION:  
; APPLICANT: Eyre, David R.  
; TITLE OF INVENTION: SYNTHETIC PEPTIDES FOR COLLAGEN RESORPTION ASSAYS  
; FILE REFERENCE: WROS-1-18220  
; CURRENT APPLICATION NUMBER: US/10/009,999A  
; CURRENT FILING DATE: 2003-01-22  
; PRIOR APPLICATION NUMBER: PCT/US99/29357  
; PRIOR FILING DATE: 1999-12-10  
; PRIOR APPLICATION NUMBER: US 09/335,098  
; PRIOR FILING DATE: 1999-06-17  
; PRIOR APPLICATION NUMBER: US 60/141,574  
; PRIOR FILING DATE: 1999-06-29  
; PRIOR APPLICATION NUMBER: US 60/142,274  
; PRIOR FILING DATE: 1999-07-02  
; PRIOR APPLICATION NUMBER: US 60/142,675  
; PRIOR FILING DATE: 1999-07-07  
; PRIOR APPLICATION NUMBER: US 09/385,740  
; PRIOR FILING DATE: 1999-08-30  
; NUMBER OF SEQ ID NOS: 45  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 42  
; LENGTH: 11  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: MISC FEATURE  
; OTHER INFORMATION: synthetic peptide corresponding to carboxy-terminal  
; OTHER INFORMATION: telopeptide sequence of human type III collagen  
US-10-009-999A-42

Query Match 100.0%; Score 42; DB 4; Length 11;  
Best Local Similarity 100.0%; Pred. No. 0.27;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 IGGEKAGG 8  
| | | | |  
Db 4 IGGEKAGG 11

RESULT 4  
US-10-009-999A-43  
; Sequence 43, Application US/10009999A  
; Patent No. 6602980  
; GENERAL INFORMATION:  
; APPLICANT: Eyre, David R.  
; TITLE OF INVENTION: SYNTHETIC PEPTIDES FOR COLLAGEN RESORPTION ASSAYS  
; FILE REFERENCE: WROS-1-18220  
; CURRENT APPLICATION NUMBER: US/10/009,999A  
; CURRENT FILING DATE: 2003-01-22  
; PRIOR APPLICATION NUMBER: PCT/US99/29357  
; PRIOR FILING DATE: 1999-12-10  
; PRIOR APPLICATION NUMBER: US 09/335,098  
; PRIOR FILING DATE: 1999-06-17  
; PRIOR APPLICATION NUMBER: US 60/141,574  
; PRIOR FILING DATE: 1999-06-29  
; PRIOR APPLICATION NUMBER: US 60/142,274  
; PRIOR FILING DATE: 1999-07-02  
; PRIOR APPLICATION NUMBER: US 60/142,675  
; PRIOR FILING DATE: 1999-07-07  
; PRIOR APPLICATION NUMBER: US 09/385,740  
; PRIOR FILING DATE: 1999-08-30  
; NUMBER OF SEQ ID NOS: 45  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 43  
; LENGTH: 12  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: MISC FEATURE  
; OTHER INFORMATION: synthetic peptide corresponding to carboxy-terminal  
; OTHER INFORMATION: telopeptide sequence of human type III collagen  
US-10-009-999A-43

Query Match 100.0%; Score 42; DB 4; Length 12;  
Best Local Similarity 100.0%; Pred. No. 0.29;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 IGGEKAGG 8  
| | | | |  
Db 4 IGGEKAGG 11

RESULT 5  
US-08-278-774-20  
; Sequence 20, Application US/08278774  
; Patent No. 6653450  
; GENERAL INFORMATION:  
; APPLICANT: Berg, Richard A  
; APPLICANT: Toman, David P  
; APPLICANT: Wallace, Donald  
; TITLE OF INVENTION: MUTATED RECOMBINANT COLLAGENS  
; NUMBER OF SEQUENCES: 33  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: COLLAGEN CORPORATION  
; STREET: 2500 Faber Place  
; CITY: Palo Alto  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94303  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/278,774

```
; FILING DATE: 22-JUL-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Rafayko, Kathi L
; REGISTRATION NUMBER: 36,644
; REFERENCE/DOCKET NUMBER: 94-018
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 354-4642
; TELEFAX: (415) 354-4752
; TELEX:
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 15 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-278-774-20

Query Match 100.0%; Score 42; DB 4; Length 15;
Best Local Similarity 100.0%; Pred. No. 0.36;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 IGGEKAGG 8
Db 2 IGGEKAGG 9

RESULT 6
US-029-348-3
; Sequence 3, Application US/09029348
; Patent No. 6171827
; GENERAL INFORMATION:
; APPLICANT: THE VICTORIA UNIVERSITY OF MANCHESTER
; TITLE OF INVENTION: NOVEL PROCOLLAGENS
; FILE REFERENCE: d087857PUS LISTING
; CURRENT APPLICATION NUMBER: US/029,348
; CURRENT FILING DATE: 1998-05-07
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 3
; LENGTH: 623
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: SEQUENCE
; OTHER INFORMATION: DERIVED FROM cDNA OF PROCOLLAGENS
US-029-348-3

Query Match 100.0%; Score 42; DB 3; Length 623;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 IGGEKAGG 8
Db 368 IGGEKAGG 375

RESULT 7
US-029-348-2
; Sequence 2, Application US/09029348
; Patent No. 6171827
; GENERAL INFORMATION:
; APPLICANT: THE VICTORIA UNIVERSITY OF MANCHESTER
; TITLE OF INVENTION: NOVEL PROCOLLAGENS
; FILE REFERENCE: d087857PUS LISTING
; CURRENT APPLICATION NUMBER: US/029,348
; CURRENT FILING DATE: 1998-05-07
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 2
; LENGTH: 626
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
```

```
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: SEQUENCE
US-029-348-2

Query Match 100.0%; Score 42; DB 3; Length 626;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 IGGEKAGG 8
Db 368 IGGEKAGG 375

RESULT 8
US-029-199-637A-51
; Sequence 51, Application US/09199637A
; Patent No. 6355411
; GENERAL INFORMATION:
; APPLICANT: Ausubel, Frederick
; APPLICANT: Goodman, Howard M.
; APPLICANT: Rahme, Laurence G.
; APPLICANT: Mahajan-Miklos, Shalina
; APPLICANT: Tan, Man-Wah
; APPLICANT: Cao, Hui
; APPLICANT: Drenkard, Eliana
; APPLICANT: Tsongalis, John
; TITLE OF INVENTION: VIRULENCE-ASSOCIATED NUCLEIC ACID
; TITLE OF INVENTION: SEQUENCES AND USES THEREOF
; FILE REFERENCE: 00786/361002
; CURRENT APPLICATION NUMBER: US/09/199,637A
; CURRENT FILING DATE: 1998-11-25
; PRIOR APPLICATION NUMBER: 60/066,517
; PRIOR FILING DATE: 1997-11-25
; NUMBER OF SEQ ID NOS: 437
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 51
; LENGTH: 674
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-029-199-637A-51

Query Match 90.5%; Score 38; DB 4; Length 674;
Best Local Similarity 75.0%; Pred. No. 67;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 IGGEKAGG 8
Db 413 VGGDKAGG 420

RESULT 9
US-0252-991A-26476
; Sequence 26476, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 26476
; LENGTH: 674
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-0252-991A-26476
```

Query Match 90.5%; Score 38; DB 4; Length 674;  
Best Local Similarity 75.0%; Pred. No. 67;  
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 IGGEKAGG 8  
Db 413 VGGDKAGG 420

RESULT 10  
5473052-20  
; Patent No. 5473052  
; APPLICANT: EYRE, DAVID R.  
; TITLE OF INVENTION: ANTIGEN-BINDING FRAGMENTS OF AN ANTIBODY  
; TO TYPE-I COLLAGEN AMINO-TERMINAL TELEPEPTIDE  
; NUMBER OF SEQUENCES: 30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/221,705  
; FILING DATE: 01-APR-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 614,719  
; FILING DATE: 21-NOV-1990  
; APPLICATION NUMBER: 444,881  
; FILING DATE: 01-DEC-1989  
; APPLICATION NUMBER: 118,234  
; FILING DATE: 06-NOV-1987  
; SEQ ID NO:20:  
; LENGTH: 11  
5473052-20

Query Match 85.7%; Score 36; DB 6; Length 11;  
Best Local Similarity 87.5%; Pred. No. 2.8;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 IGGEKAGG 8  
Db 2 IGGEKAGG 9

RESULT 11  
US-08-963-825-21  
; Sequence 21, Application US/08963825  
; Patent No. 6110689  
; GENERAL INFORMATION:  
; APPLICANT: Qvist, Per  
; APPLICANT: Bonde, Martin  
; TITLE OF INVENTION: A Method for Assaying Collagen Fragments  
; TITLE OF INVENTION: in Body Fluids, A Test Kit and Means for Carrying Out the  
; TITLE OF INVENTION: Method and Use of the Method to Diagnose the Presence of  
; TITLE OF INVENTION: Disorders Associated with the Metabolism of  
; NUMBER OF SEQUENCES: 21  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Darby & Darby PC  
; STREET: 805 Third Avenue  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10022  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/963,825  
; FILING DATE:  
; CLASSIFICATION: 436  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/08/187,319  
; FILING DATE: 21-JAN-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Gogoris, Adda C

REGISTRATION NUMBER: 29,714  
REFERENCE/DOCKET NUMBER: 4305/08701  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-527-7700  
TELEFAX: 212-753-6237  
TELEX: 236687  
INFORMATION FOR SEQ ID NO: 21:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1078 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
ORIGINAL SOURCE:  
ORGANISM: Homo sapiens  
IMMEDIATE SOURCE:  
CLONE: COLLAGEN ALPHA 1 (III)  
US-08-963-825-21

Query Match 85.7%; Score 36; DB 3; Length 1078;  
Best Local Similarity 87.5%; Pred. No. 2.3e+02;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 IGGEKAGG 8  
Db 1061 IGAEKAGG 1068

RESULT 12  
US-09-500-811-21  
; Sequence 21, Application US/09500811  
; Patent No. 6323314  
; GENERAL INFORMATION:  
; APPLICANT: Qvist, Per  
; APPLICANT: Bonde, Martin  
; TITLE OF INVENTION: A Method for Assaying Collagen Fragments  
; TITLE OF INVENTION: in Body Fluids, A Test Kit and Means for Carrying Out the  
; TITLE OF INVENTION: Method and Use of the Method to Diagnose the Presence of  
; TITLE OF INVENTION: Disorders Associated with the Metabolism of  
; NUMBER OF SEQUENCES: 21  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Darby & Darby PC  
; STREET: 805 Third Avenue  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10022  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/500,811  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/187,319  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Gogoris, Adda C  
; REGISTRATION NUMBER: 29,714  
; REFERENCE/DOCKET NUMBER: 4305/08701  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212-527-7700  
; TELEFAX: 212-753-6237  
; TELEX: 236687  
; INFORMATION FOR SEQ ID NO: 21:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1078 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
ORIGINAL SOURCE:

```
; ORGANISM: Homo sapiens
; IMMEDIATE SOURCE:
; CLONE: COLLAGEN ALPHA 1 (III)
US-09-500-811-21
Query Match      85.7%; Score 36; DB 4; Length 1078;
Best Local Similarity 87.5%; Pred. No. 2.3e+02;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 IGGEKAGG 8
|||
Db 1061 IGAEKAGG 1068

RESULT 13
US-09-570-573-21
; Sequence 21, Application US/09570573
; Patent No. 6342361
; GENERAL INFORMATION:
; APPLICANT: Qvist, Per
; APPLICANT: Bonde, Martin
; TITLE OF INVENTION: A Method for Assaying Collagen Fragments
; TITLE OF INVENTION: in Body Fluids, A Test Kit and Means for Carrying Out the
; TITLE OF INVENTION: Method and Use of the Method to Diagnose the Presence of
; TITLE OF INVENTION: Disorders Associated with the Metabolism of
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Darby & Darby PC
; STREET: 805 Third Avenue
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/570,573
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/187,319
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Gogoris, Adda C
; REGISTRATION NUMBER: 29,714
; REFERENCE/DOCKET NUMBER: 4305/08701
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-527-7700
; TELEFAX: 212-753-6237
; TELEX: 236687
; INFORMATION FOR SEQ ID NO: 21:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1078 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; IMMEDIATE SOURCE:
; CLONE: COLLAGEN ALPHA 1 (III)
US-09-570-573-21
Query Match      85.7%; Score 36; DB 4; Length 1078;
Best Local Similarity 87.5%; Pred. No. 2.3e+02;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 IGGEKAGG 8
|||
Db 1061 IGAEKAGG 1068

RESULT 14
US-09-548-608-21
; Sequence 21, Application US/09548608
; Patent No. 6355442
; GENERAL INFORMATION:
; APPLICANT: Qvist, Per
; APPLICANT: Bonde, Martin
; TITLE OF INVENTION: A Method for Assaying Collagen Fragments
; TITLE OF INVENTION: in Body Fluids, A Test Kit and Means for Carrying Out the
; TITLE OF INVENTION: Method and Use of the Method to Diagnose the Presence of
; TITLE OF INVENTION: Disorders Associated with the Metabolism of
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Darby & Darby PC
; STREET: 805 Third Avenue
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/548,608
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/187,319
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Gogoris, Adda C
; REGISTRATION NUMBER: 29,714
; REFERENCE/DOCKET NUMBER: 4305/08701
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-527-7700
; TELEFAX: 212-753-6237
; TELEX: 236687
; INFORMATION FOR SEQ ID NO: 21:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1078 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; IMMEDIATE SOURCE:
; CLONE: COLLAGEN ALPHA 1 (III)
US-09-548-608-21
Query Match      85.7%; Score 36; DB 4; Length 1078;
Best Local Similarity 87.5%; Pred. No. 2.3e+02;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 IGGEKAGG 8
|||
Db 1061 IGAEKAGG 1068

RESULT 15
US-09-252-991A-28074
; Sequence 28074, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
```

Mon Sep 20 11:05:43 2004

```
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 28074
; LENGTH: 561
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-28074

Query Match      83.3%; Score 35; DB 4; Length 561;
Best Local Similarity 75.0%; Pred. No. 1.8e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY      1 IGGEKAGG 8
      :|||
Db      63 VGGEDAGG 70

Search completed: September 18, 2004, 04:32:21
Job time : 14.1765 secs
```



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 18, 2004, 04:29:56 ; Search time 37.9608 Seconds  
(without alignments)  
67.677 Million cell updates/sec

Title: US-10-615-959-40  
Perfect score: 42  
Sequence: 1 IGGERAGG 8

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1342398 seqs, 32113274 residues

Total number of hits satisfying chosen parameters: 1342398

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications AA:\*

1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*  
2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*  
8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/2/pubpaa/US09A\_PUBCOMB.pep.\*  
10: /cgn2\_6/ptodata/2/pubpaa/US09B\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/2/pubpaa/US09C\_PUBCOMB.pep.\*  
12: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep.\*  
13: /cgn2\_6/ptodata/2/pubpaa/US10A\_PUBCOMB.pep.\*  
14: /cgn2\_6/ptodata/2/pubpaa/US10B\_PUBCOMB.pep.\*  
15: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pep.\*  
16: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep.\*  
17: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*  
18: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	42	100.0	8	12	US-10-615-959-40
2	42	100.0	9	12	US-10-615-959-41
3	42	100.0	11	12	US-10-615-959-42
4	42	100.0	12	12	US-10-615-959-43
5	42	100.0	25	15	US-10-366-125-11
6	42	100.0	262	15	US-10-366-125-19
7	42	100.0	309	9	US-09-908-711-78
8	42	100.0	1466	12	US-09-918-715-226
9	42	100.0	1466	14	US-10-257-021-72
10	42	100.0	1466	14	US-10-177-293-68
11	42	100.0	1466	14	US-10-301-822-33
12	42	100.0	1466	16	US-10-357-851-3
13	42	100.0	1466	16	US-10-358-024-3
14	42	100.0	1466	16	US-10-734-564-103
15	41	97.6	1466	15	US-10-402-089-12

16	41	97.6	1466	15	US-10-402-072A-12	Sequence 12, Appl
17	38	90.5	322	16	US-10-437-963-137654	Sequence 137654,
18	38	90.5	674	10	US-09-975-719-51	Sequence 51, Appl
19	38	90.5	696	16	US-10-437-963-189506	Sequence 189506,
20	37	88.1	225	16	US-10-437-963-106003	Sequence 106003,
21	37	88.1	969	16	US-10-437-963-165327	Sequence 165327,
22	36	85.7	73	12	US-10-424-599-249081	Sequence 249081,
23	36	85.7	155	12	US-10-424-599-190094	Sequence 190094,
24	36	85.7	353	14	US-10-156-761-7849	Sequence 7849, Ap
25	36	85.7	1078	14	US-10-058-124-21	Sequence 21, Appl
26	35	83.3	53	12	US-10-424-599-279107	Sequence 279107,
27	35	83.3	60	16	US-10-437-963-189501	Sequence 189501,
28	35	83.3	172	12	US-10-425-114-61090	Sequence 61090, A
29	35	83.3	249	16	US-10-437-963-157383	Sequence 157383,
30	35	83.3	320	14	US-10-214-473-6	Sequence 6, Appli
31	35	83.3	320	14	US-10-272-490-6	Sequence 6, Appli
32	35	83.3	457	15	US-10-369-493-18766	Sequence 18766, A
33	35	83.3	467	15	US-10-369-493-19844	Sequence 19844, A
34	35	83.3	476	15	US-10-369-493-21046	Sequence 21046, A
35	35	83.3	490	15	US-10-369-493-2735	Sequence 2735, Ap
36	35	83.3	516	16	US-10-437-963-188076	Sequence 188076,
37	35	83.3	691	12	US-10-282-122A-47568	Sequence 47568, A
38	35	83.3	1466	15	US-10-402-089-4	Sequence 4, Appli
39	35	83.3	1466	15	US-10-402-089-6	Sequence 6, Appli
40	35	83.3	1466	15	US-10-402-072A-4	Sequence 4, Appli
41	35	83.3	1466	15	US-10-402-072A-6	Sequence 6, Appli
42	34	81.0	66	9	US-09-864-761-38038	Sequence 38038, A
43	34	81.0	75	12	US-10-424-599-185108	Sequence 185108,
44	34	81.0	80	16	US-10-437-963-196331	Sequence 196331,
45	34	81.0	98	16	US-10-437-963-107817	Sequence 107817,

## ALIGNMENTS

## RESULT 1

US-10-615-959-40  
; Sequence 40, Application US/10615959  
; Publication No. US20040048321A1  
; GENERAL INFORMATION:  
; APPLICANT: Eyre, David R.  
; TITLE OF INVENTION: SYNTHETIC PEPTIDES FOR COLLAGEN RESORPTION ASSAYS  
; FILE REFERENCE: WROS-1-18220  
; CURRENT APPLICATION NUMBER: US/10/615,959  
; CURRENT FILING DATE: 2003-07-08  
; PRIOR APPLICATION NUMBER: US/10/009,999A  
; PRIOR FILING DATE: 2003-01-22  
; PRIOR APPLICATION NUMBER: PCT/US99/29357  
; PRIOR FILING DATE: 1999-12-10  
; PRIOR APPLICATION NUMBER: US 09/335,098  
; PRIOR FILING DATE: 1999-06-17  
; PRIOR APPLICATION NUMBER: US 60/141,574  
; PRIOR FILING DATE: 1999-06-29  
; PRIOR APPLICATION NUMBER: US 60/142,274  
; PRIOR FILING DATE: 1999-07-02  
; PRIOR APPLICATION NUMBER: US 60/142,675  
; PRIOR FILING DATE: 1999-07-07  
; PRIOR APPLICATION NUMBER: US 09/385,740  
; PRIOR FILING DATE: 1999-08-30  
; NUMBER OF SEQ ID NOS: 45  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 40  
; LENGTH: 8  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: MISC\_FEATURE  
; OTHER INFORMATION: synthetic peptide corresponding to carboxy-terminal  
; OTHER INFORMATION: telopeptide sequence of human type III collagen  
US-10-615-959-40

Query Match 100.0%; Score 42; DB 12; Length 8;  
Best Local Similarity 100.0%; Pred. No. 1.2e+06;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IGGEKAGG 8  
| | | | |  
Db 1 IGGEKAGG 8

RESULT 2

US-10-615-959-41  
; Sequence 41, Application US/10615959  
; Publication No. US20040048321A1  
; GENERAL INFORMATION:  
; APPLICANT: Eyre, David R.  
; TITLE OF INVENTION: SYNTHETIC PEPTIDES FOR COLLAGEN RESORPTION ASSAYS  
; FILE REFERENCE: WROS-1-18220  
; CURRENT APPLICATION NUMBER: US/10/615,959  
; CURRENT FILING DATE: 2003-07-08  
; PRIOR APPLICATION NUMBER: US/10/009,999A  
; PRIOR FILING DATE: 2003-01-22  
; PRIOR APPLICATION NUMBER: PCT/US99/29357  
; PRIOR FILING DATE: 1999-12-10  
; PRIOR APPLICATION NUMBER: US 09/335,098  
; PRIOR FILING DATE: 1999-06-17  
; PRIOR APPLICATION NUMBER: US 60/141,574  
; PRIOR FILING DATE: 1999-06-29  
; PRIOR APPLICATION NUMBER: US 60/142,274  
; PRIOR FILING DATE: 1999-07-02  
; PRIOR APPLICATION NUMBER: US 60/142,675  
; PRIOR FILING DATE: 1999-07-07  
; PRIOR APPLICATION NUMBER: US 09/385,740  
; PRIOR FILING DATE: 1999-08-30  
; NUMBER OF SEQ ID NOS: 45  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 41  
; LENGTH: 9  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; NAME/KEY: MISC FEATURE  
; OTHER INFORMATION: synthetic peptide corresponding to carboxy-terminal  
; OTHER INFORMATION: telopeptide sequence of human type III collagen  
US-10-615-959-41

Query Match 100.0%; Score 42; DB 12; Length 9;  
Best Local Similarity 100.0%; Pred. No. 1.2e+06;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IGGEKAGG 8  
| | | | |  
Db 1 IGGEKAGG 8

RESULT 3

US-10-615-959-42  
; Sequence 42, Application US/10615959  
; Publication No. US20040048321A1  
; GENERAL INFORMATION:  
; APPLICANT: Eyre, David R.  
; TITLE OF INVENTION: SYNTHETIC PEPTIDES FOR COLLAGEN RESORPTION ASSAYS  
; FILE REFERENCE: WROS-1-18220  
; CURRENT APPLICATION NUMBER: US/10/615,959  
; CURRENT FILING DATE: 2003-07-08  
; PRIOR APPLICATION NUMBER: US/10/009,999A  
; PRIOR FILING DATE: 2003-01-22  
; PRIOR APPLICATION NUMBER: PCT/US99/29357  
; PRIOR FILING DATE: 1999-12-10  
; PRIOR APPLICATION NUMBER: US 09/335,098  
; PRIOR FILING DATE: 1999-06-17  
; PRIOR APPLICATION NUMBER: US 60/141,574  
; PRIOR FILING DATE: 1999-06-29  
; PRIOR APPLICATION NUMBER: US 60/142,274  
; PRIOR FILING DATE: 1999-07-02  
; PRIOR APPLICATION NUMBER: US 60/142,675

; PRIOR FILING DATE: 1999-07-07  
; PRIOR APPLICATION NUMBER: US 09/385,740  
; PRIOR FILING DATE: 1999-08-30  
; NUMBER OF SEQ ID NOS: 45  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 42  
; LENGTH: 11  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; NAME/KEY: MISC FEATURE  
; OTHER INFORMATION: synthetic peptide corresponding to carboxy-terminal  
; OTHER INFORMATION: telopeptide sequence of human type III collagen  
US-10-615-959-42

Query Match 100.0%; Score 42; DB 12; Length 11;  
Best Local Similarity 100.0%; Pred. No. 1.6;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IGGEKAGG 8  
| | | | |  
Db 4 IGGEKAGG 11

RESULT 4

US-10-615-959-43  
; Sequence 43, Application US/10615959  
; Publication No. US20040048321A1  
; GENERAL INFORMATION:  
; APPLICANT: Eyre, David R.  
; TITLE OF INVENTION: SYNTHETIC PEPTIDES FOR COLLAGEN RESORPTION ASSAYS  
; FILE REFERENCE: WROS-1-18220  
; CURRENT APPLICATION NUMBER: US/10/615,959  
; CURRENT FILING DATE: 2003-07-08  
; PRIOR APPLICATION NUMBER: US/10/009,999A  
; PRIOR FILING DATE: 2003-01-22  
; PRIOR APPLICATION NUMBER: PCT/US99/29357  
; PRIOR FILING DATE: 1999-12-10  
; PRIOR APPLICATION NUMBER: US 09/335,098  
; PRIOR FILING DATE: 1999-06-17  
; PRIOR APPLICATION NUMBER: US 60/141,574  
; PRIOR FILING DATE: 1999-06-29  
; PRIOR APPLICATION NUMBER: US 60/142,274  
; PRIOR FILING DATE: 1999-07-02  
; PRIOR APPLICATION NUMBER: US 60/142,675  
; PRIOR FILING DATE: 1999-07-07  
; PRIOR APPLICATION NUMBER: US 09/385,740  
; PRIOR FILING DATE: 1999-08-30  
; NUMBER OF SEQ ID NOS: 45  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 43  
; LENGTH: 12  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; NAME/KEY: MISC FEATURE  
; OTHER INFORMATION: synthetic peptide corresponding to carboxy-terminal  
; OTHER INFORMATION: telopeptide sequence of human type III collagen  
US-10-615-959-43

Query Match 100.0%; Score 42; DB 12; Length 12;  
Best Local Similarity 100.0%; Pred. No. 1.8;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IGGEKAGG 8  
| | | | |  
Db 4 IGGEKAGG 11

RESULT 5

US-10-366-125-11  
; Sequence 11, Application US/10366125  
; Publication No. US20030228259A1

```

; GENERAL INFORMATION:
; APPLICANT: Hellerstein, Marc
; TITLE OF INVENTION: MEASUREMENT OF BIOSYNTHESIS AND BREAKDOWN RATES OF
; TITLE OF INVENTION: BIOLOGICAL MOLECULES THAT ARE INACCESSIBLE OR NOT
; TITLE OF INVENTION: EASILY ACCESSIBLE TO DIRECT SAMPLING, NON-INVASIVELY,
; TITLE OF INVENTION: BY LABEL INCORPORATION INTO METABOLIC DERIVATIVES AND
; TITLE OF INVENTION: CATABOLITIC PRODUCTS
; FILE REFERENCE: 416272003500
; CURRENT APPLICATION NUMBER: US/10/366,125
; CURRENT FILING DATE: 2003-02-12
; PRIOR APPLICATION NUMBER: US 60/356,008
; PRIOR FILING DATE: 2002-02-12
; NUMBER OF SEQ ID NOS: 28
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 25
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-366-125-11

Query Match      100.0%; Score 42; DB 15; Length 25;
Best Local Similarity 100.0%; Pred. No. 3.6;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 IGGEKAGG 8
      |||||
Db      12 IGGEKAGG 19

RESULT 6
US-10-366-125-19
; Sequence 19, Application US/10366125
; Publication No. US20030228259A1
; GENERAL INFORMATION:
; APPLICANT: Hellerstein, Marc
; TITLE OF INVENTION: MEASUREMENT OF BIOSYNTHESIS AND BREAKDOWN RATES OF
; TITLE OF INVENTION: BIOLOGICAL MOLECULES THAT ARE INACCESSIBLE OR NOT
; TITLE OF INVENTION: EASILY ACCESSIBLE TO DIRECT SAMPLING, NON-INVASIVELY,
; TITLE OF INVENTION: BY LABEL INCORPORATION INTO METABOLIC DERIVATIVES AND
; TITLE OF INVENTION: CATABOLITIC PRODUCTS
; FILE REFERENCE: 416272003500
; CURRENT APPLICATION NUMBER: US/10/366,125
; CURRENT FILING DATE: 2003-02-12
; PRIOR APPLICATION NUMBER: US 60/356,008
; PRIOR FILING DATE: 2002-02-12
; NUMBER OF SEQ ID NOS: 28
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 19
; LENGTH: 262
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-366-125-19

Query Match      100.0%; Score 42; DB 15; Length 262;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 IGGEKAGG 8
      |||||
Db      4 IGGEKAGG 11

RESULT 7
US-09-908-711-78
; Sequence 78, Application US/09908711
; Patent No. US20020045230A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PA128
; CURRENT APPLICATION NUMBER: US/09/908,711
; CURRENT FILING DATE: 2001-07-20
; PRIOR APPLICATION NUMBER: US01/01360
; NUMBER OF SEQ ID NOS: 167
; SOFTWARE: PatentIn ver. 2.0
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,867
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01344
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,892
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01345
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,888
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01329
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,905
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01354
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,891
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01339
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,869
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01340
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,874
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01334
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,898
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01320
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,853
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01349
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,902
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01239
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,870
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01348
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,882
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01347
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,896
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01307
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,864
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01341
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,856
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01336
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,868
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01312
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 60/179,065
; PRIOR FILING DATE: 2000-01-31
; PRIOR APPLICATION NUMBER: 60/180,628
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: 60/209,467
; PRIOR FILING DATE: 2000-06-07
; NUMBER OF SEQ ID NOS: 167
; SOFTWARE: PatentIn ver. 2.0

```

Mon Sep 20 11:05:43 2004

```
; SEQ ID NO 78
; LENGTH: 309
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (2)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (4)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-908-711-78
```

```
Query Match 100.0%; Score 42; DB 9; Length 309;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 1 IGGEKAGG 8
Db 51 IGGEKAGG 58
```

```
RESULT 8
US-09-918-715-226
; Sequence 226, Application US/09918715
; Publication No. US20030017157A1
; GENERAL INFORMATION:
; APPLICANT: Brad St. Croix
; APPLICANT: Bert Vogelstein
; APPLICANT: Kenneth Kinzler
; TITLE OF INVENTION: ENDOTHELIAL CELL EXPRESSION PATTERNS
; FILE REFERENCE: 1107.00134
; CURRENT APPLICATION NUMBER: US/09/918,715
; CURRENT FILING DATE: 2001-08-01
; PRIOR APPLICATION NUMBER: 60/222,599
; PRIOR FILING DATE: 2000-08-02
; PRIOR APPLICATION NUMBER: 60/224,360
; PRIOR FILING DATE: 2000-08-11
; PRIOR APPLICATION NUMBER: 60/282,850
; PRIOR FILING DATE: 2000-04-11
; NUMBER OF SEQ ID NOS: 358
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 226
; LENGTH: 1466
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-918-715-226
```

```
Query Match 100.0%; Score 42; DB 12; Length 1466;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 1 IGGEKAGG 8
Db 1208 IGGEKAGG 1215
```

```
RESULT 9
US-10-257-021-72
; Sequence 72, Application US/10257021
; Publication No. US20030211498A1
; GENERAL INFORMATION:
; APPLICANT: Morin, Patrice J.
; APPLICANT: Sherman-Baust, Cheryl A.
; APPLICANT: Pizer, Ellen S.
; APPLICANT: Hough, Colleen D.
; TITLE OF INVENTION: TUMOR MARKERS IN OVARIAN CANCER
; FILE REFERENCE: 14014.0369U2
; CURRENT APPLICATION NUMBER: US/10/257,021
; CURRENT FILING DATE: 2002-10-03
; PRIOR APPLICATION NUMBER: PCT/US01/10947
; PRIOR FILING DATE: 2001-04-03
; PRIOR APPLICATION NUMBER: 60/194,336
```

```
; PRIOR FILING DATE: 2000-04-03
; NUMBER OF SEQ ID NOS: 147
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 72
; LENGTH: 1466
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-257-021-72
```

```
Query Match 100.0%; Score 42; DB 12; Length 1466;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 1 IGGEKAGG 8
Db 1208 IGGEKAGG 1215
```

```
RESULT 10
US-10-177-293-68
; Sequence 68, Application US/10177293
; Publication No. US20030124128A1
; GENERAL INFORMATION:
; APPLICANT: Lillie, Karen
; APPLICANT: Zhao, Xumei
; APPLICANT: Gannavarpu, Manjula
; APPLICANT: Kamatkar, Shubhangi
; APPLICANT: Mertens, Maureen
; APPLICANT: Myer, Vic
; APPLICANT: Wang, Youzhen
; APPLICANT: Xu, Yongyao
; APPLICANT: Hoersch, Sebastian
; APPLICANT: Monahan, John
; APPLICANT: Meyers, Rachel E.
; APPLICANT: Bast Jr., Robert C.
; APPLICANT: Hortobagyi, Gabriel N.
; APPLICANT: Pusztai Lajos
; APPLICANT: Meric, Funda
; APPLICANT: Sahin, Aysegul
; APPLICANT: Mills, Gordon B.
; TITLE OF INVENTION: COMPOSITIONS, KITS, AND METHODS FOR IDENTIFICATION, ASSESSMENT,
; FILE REFERENCE: MEI-038
; CURRENT APPLICATION NUMBER: US/10/177,293
; CURRENT FILING DATE: 2002-06-21
; PRIOR APPLICATION NUMBER: US 60/299,887
; PRIOR FILING DATE: 2001-06-21
; PRIOR APPLICATION NUMBER: US 60/301,572
; PRIOR FILING DATE: 2001-06-27
; PRIOR APPLICATION NUMBER: US 60/306,501
; PRIOR FILING DATE: 2001-07-18
; PRIOR APPLICATION NUMBER: US 60/325,002
; PRIOR FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: US 60/362,585
; PRIOR FILING DATE: 2002-03-05
; PRIOR APPLICATION NUMBER: US 60/xxx,xxx
; PRIOR FILING DATE: 2002-05-14
; NUMBER OF SEQ ID NOS: 506
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 68
; LENGTH: 1466
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-177-293-68
```

```
Query Match 100.0%; Score 42; DB 14; Length 1466;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 1 IGGEKAGG 8
Db 1208 IGGEKAGG 1215
```

```
RESULT 11
US-10-301-822-33
; Sequence 33, Application US/10301822
; Publication No. US20030148410A1
; GENERAL INFORMATION:
; APPLICANT: Millennium Pharmaceuticals, Inc.
; APPLICANT: Berger, Allison
; APPLICANT: Guillemette, Tracy L.
; APPLICANT: Kamatkar, Shubhangi
; APPLICANT: Schlegel, Robert
; APPLICANT: Monahan, John E.
; APPLICANT: Thibodeau, Stephen N.
; APPLICANT: Burgart, Lawrence J.
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND
; TITLE OF INVENTION: METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF COLON CANCER
; FILE REFERENCE: MP001-029P2RNM
; CURRENT APPLICATION NUMBER: US/10/301,822
; CURRENT FILING DATE: 2002-11-21
; PRIOR APPLICATION NUMBER: US 60/339,971
; PRIOR FILING DATE: 2001-12-10
; PRIOR APPLICATION NUMBER: US 60/361,978
; PRIOR FILING DATE: 2002-03-05
; PRIOR APPLICATION NUMBER: US 60/381,988
; PRIOR FILING DATE: 2002-05-20
; NUMBER OF SEQ ID NOS: 228
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 33
; LENGTH: 1466
; TYPE: PRT
; ORGANISM: Homo Sapiens
US-10-301-822-33

Query Match      100.0%; Score 42; DB 14; Length 1466;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 IGGEKAGG 8
      |||||
Db      1208 IGGEKAGG 1215

RESULT 12
US-10-357-851-3
; Sequence 3, Application US/10357851
; Publication No. US20040151731A1
; GENERAL INFORMATION:
; APPLICANT: Jicha, Douglas L.
; TITLE OF INVENTION: Method and Compositions Involving
; TITLE OF INVENTION: T-Lymphocyte Reactivity with Collagen in Aortic Tissue of
; TITLE OF INVENTION: Abdominal Aortic Aneurysm Patients
; FILE REFERENCE: 13376US
; CURRENT APPLICATION NUMBER: US/10/357,851
; CURRENT FILING DATE: 2003-02-04
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 1466
; TYPE: PRT
; ORGANISM: Homo Sapiens
US-10-357-851-3

Query Match      100.0%; Score 42; DB 16; Length 1466;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 IGGEKAGG 8
      |||||
Db      1208 IGGEKAGG 1215

RESULT 13
US-10-358-024-3
; Sequence 3, Application US/10358024
; Publication No. US20040151732A1
; GENERAL INFORMATION:
; APPLICANT: Jicha, Douglas L.
; APPLICANT: Pelsue, Stephen
; TITLE OF INVENTION: Methods and Compositions Involving Blood
; TITLE OF INVENTION: T-Lymphocytes Reactivity with Collagen
; FILE REFERENCE: 13436US
; CURRENT APPLICATION NUMBER: US/10/358,024
; CURRENT FILING DATE: 2003-02-04
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 1466
; TYPE: PRT
; ORGANISM: Homo Sapiens
US-10-358-024-3

Query Match      100.0%; Score 42; DB 16; Length 1466;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 IGGEKAGG 8
      |||||
Db      1208 IGGEKAGG 1215

RESULT 14
US-10-734-564-103
; Sequence 103, Application US/10734564
; Publication No. US20040157278A1
; GENERAL INFORMATION:
; APPLICANT: Christopher C Burgess et al
; TITLE OF INVENTION: Detection Methods Using TIMP1
; FILE REFERENCE: 1657/2012
; CURRENT APPLICATION NUMBER: US/10/734,564
; CURRENT FILING DATE: 2003-12-12
; NUMBER OF SEQ ID NOS: 138
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 103
; LENGTH: 1466
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-734-564-103

Query Match      100.0%; Score 42; DB 16; Length 1466;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 IGGEKAGG 8
      |||||
Db      1208 IGGEKAGG 1215

RESULT 15
US-10-402-089-12
; Sequence 12, Application US/10402089
; Publication No. US20040005663A1
; GENERAL INFORMATION:
; APPLICANT: Bell, Marcum P.
; APPLICANT: Neff, Thomas B.
; APPLICANT: Polarek, James W.
; APPLICANT: Seeley, Todd W.
; TITLE OF INVENTION: PORCINE COLLAGENS AND GELATINS
; FILE REFERENCE: F00402.3 CON
; CURRENT APPLICATION NUMBER: US/10/402,089
; CURRENT FILING DATE: 2003-03-26
; PRIOR APPLICATION NUMBER: US 09/709,700
; PRIOR FILING DATE: 2000-11-10
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: PatentIn version 3.2
```

Mon Sep 20 11:05:43 2004

```
; SEQ ID NO 12
; LENGTH: 1466
; TYPE: PRT
; ORGANISM: Sus scrofa
US-10-402-089-12

Query Match      97.8%; Score 41; DB 15; Length 1466;
Best Local Similarity 87.5%; Pred. No. 2.6e+02;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy      1 IGGKKAGG 8
      :|||||
Db      1208 VGGKKAGG 1215

Search completed: September 18, 2004, 04:56:12
Job time : 38.9608 secs
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 18, 2004, 04:02:05 ; Search time 9.72549 Seconds  
(without alignments)  
79.125 Million cell updates/sec

Title: US-10-615-959-40

Perfect score: 42

Sequence: 1 IGGEKAGG 8

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR\_78.\*

1: pir1.\*

2: pir2.\*

3: pir3.\*

4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	42	100.0	1466	1 CGHU7L	collagen alpha 1(I)
2	38	90.5	636	2 S41067	collagen alpha 1(I)
3	38	90.5	1464	2 S59856	collagen alpha 1(I)
4	37	88.1	476	2 AG1031	hypothetical prote
5	36	85.7	404	2 S56535	hypothetical 43.7K
6	36	85.7	404	2 E91287	hypothetical prote
7	36	85.7	404	2 H86128	hypothetical prote
8	35	83.3	161	2 S14998	heat shock protein
9	35	83.3	270	2 H69074	formylmethanofuran
10	35	83.3	270	2 S57457	phosphoglucosylase
11	35	83.3	475	2 A52301	phosphoglucosylase
12	35	83.3	490	2 S75539	hypothetical prote
13	34	81.0	165	2 T41103	60S ribosomal prot
14	34	81.0	240	2 B36724	hypothetical prote
15	34	81.0	266	2 T31264	cis-1,2-dihydro-1,
16	34	81.0	470	2 F82302	probable phosphogl
17	34	81.0	1137	1 WMBE81	ribonucleoside-dip
18	33	78.6	113	2 T36039	ABC excision nucle
19	33	78.6	163	2 T47394	hypothetical prote
20	33	78.6	295	2 A00306	probable membrane
21	33	78.6	311	2 C59464	carboxylesterase (
22	33	78.6	364	2 I39567	probable lipase (E
23	33	78.6	376	2 G70671	probable lipase/es
24	33	78.6	383	2 E96972	riboflavin biosynt
25	33	78.6	478	2 A12441	phosphoglucosylase
26	33	78.6	773	2 T46010	anthranilate phosph
27	32	76.2	126	2 A11951	hypothetical prote
28	32	76.2	155	2 S43775	desiccation-relate
29	32	76.2	209	2 A52981	arylesterase [impo

## ALIGNMENTS

### RESULT 1

CGHU7L

collagen alpha 1(III) chain precursor - human

N:Alternate names: procollagen alpha 1(III) chain

C:Species: Homo sapiens (man)

C>Date: 24-Apr-1984 #sequence revision 01-Sep-1995 #text change 21-Jul-2000

C:Accession: S05272; S04642; PE0011; S01726; S04887; A90399; A94562; I51868; S59511; A904

R:Prockop, D.J.

submitted to the EMBL Data Library, February 1989

A:Reference number: S05272

A:Accession: S05272

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-1240,'V',1242-1466 <PRC>

A:Cross-references: EMBL:X14420; NID:G30057; PIDN:CAA32583.1; PID:G30058

R:Ala-Kokko, L.; Kontusaari, S.; Baldwin, C.T.; Kuivaniemi, H.; Prockop, D.J.

Biochem. J. 260, 509-516, 1989

A>Title: Structure of cDNA clones coding for the entire prepro-alpha(III) chain of human

erences.

A:Reference number: S04642; MUID:89350838; PMID:2764886

A:Accession: S04642

A:Molecule type: mRNA

A:Residues: 1-1196 <ALAA>

A:Cross-references: EMBL:X14420; NID:G30057; PIDN:CAA32583.1; PID:G30058

A>Note: the complete sequence is not shown

R:Benson-Chanda, V.; Su, M.W.; Weil, D.; Chu, M.L.; Ramirez, F.

Gene 78, 255-265, 1989

A>Title: Cloning and analysis of the 5' portion of the human type-III procollagen gene (

A:Reference number: PE0011; MUID:89378752; PMID:2777083

A:Accession: PE0011

A:Molecule type: DNA

A:Residues: 1-176 <BEN>

A:Cross-references: GB:M26939; NID:G180813; PIDN:AAA52040.1; PID:G180814

R:Toman, P.D.; Ricca, G.A.; de Crombrughe, B.

Nucleic Acids Res. 16, 7201, 1988

A>Title: Nucleotide sequence of a cDNA coding for the amino-terminal region of human prepr

A:Reference number: S01726; MUID:88303360; PMID:3405773

A:Accession: S01726

A:Molecule type: mRNA

A:Cross-references: EMBL:X07240; NID:G30060; PIDN:CAA30229.1; PID:G30061

A>Note: the authors translated the codon CAG for residue 154 as His

R:Janeczko, R.A.; Ramirez, F.

Nucleic Acids Res. 17, 6742, 1989

A>Title: Nucleotide and amino acid sequences of the entire human alpha-1 (III) collagen.

A:Reference number: S04887; MUID:89386015; PMID:2780304

A:Accession: S04887

A:Molecule type: mRNA

A:Residues: 149-163,'G',164-240,'D',242-471,'D',473-487,'L',489,'S',491-613,'Y',615-634,'

A:Cross-references: EMBL:X15332; NID:G29545; PIDN:CAA33387.1; PID:G930045

A>Note: the authors' translation of residues 905-932 is inconsistent with the nucleotide

R:Seyer, J.M.; Kang, A.H.

Mon Sep 20 11:05:44 2004

Biochemistry 16, 1158-1164, 1977  
 A>Title: Covalent structure of collagen: amino acid sequence of cyanogen bromide peptide  
 A;Residues: 950-1018, 'Y', 1020-1183, 'S', 1185-1466 <MAN>  
 A;Cross-references: EMBL:X06700; NID:g30053; PIDN:CAA29886.1; PID:g30054  
 A;Accession: A90399  
 A;Molecule type: protein  
 A;Residues: 'Y', 169-225, 229-232, 'P', 234-292, 'D', 294-398 <SEY1>  
 A;Experimental source: liver  
 A;Note: sequence corrected by A94562; attachment of 2-O-alpha-D-glucosyl-O-beta-D-galact  
 R;Seyer, J.M.  
 submitted to the Atlas, December 1977  
 A;Reference number: A94562  
 A;Accession: A94562  
 A;Molecule type: protein  
 A;Residues: 'V', 169-225, 229-277, 'A', 279-292, 'D', 294, 'S', 296-398 <SEY2>  
 A;Experimental source: liver  
 A;Note: author submitted corrections to A90399  
 R;Milewicz, D.M.; Witz, A.M.; Smith, A.C.; Manchester, D.K.; Waldstein, G.; Byers, P.H.  
 Am J. Hum. Genet. 53, 62-70, 1993  
 A>Title: Parental somatic and germ-line mosaicism for a multiexon deletion with unusual  
 ispring.  
 A;Reference number: I51868; MUID:93304430; PMID:8317500  
 A;Accession: I51868  
 A;Status: preliminary; translated from GB/EMBL/DBJ  
 A;Molecule type: DNA  
 A;Residues: 186-194 <MI>  
 A;Cross-references: GB:S62925; NID:g386425; PIDN:AA13937.1; PID:g4261637  
 R;Chiodo, A.A.; Sillence, D.O.; Cole, W.G.; Bateman, J.F.  
 Biochem. J. 311, 939-943, 1995  
 A>Title: Abnormal type III collagen produced by an exon-17-skipping mutation of the COL3  
 A;Reference number: S59511; MUID:96067614; PMID:7487954  
 A;Accession: S59511  
 A;Molecule type: mRNA  
 A;Residues: 302-423 <CHI>  
 A;Cross-references: GB:S79877; NID:g1195576; PIDN:AAB35615.1; PID:g1195577  
 R;Seyer, J.M.; Kang, A.H.  
 Biochemistry 17, 3404-3411, 1978  
 A>Title: Covalent structure of collagen: amino acid sequence of five consecutive CNBr pe  
 A;Reference number: A90414; MUID:79000343; PMID:687591  
 A;Accession: A90414  
 A;Molecule type: protein  
 A;Residues: 399-675, 'N', 677-727 <SEY3>  
 A;Experimental source: liver  
 R;Lee, B.; Vitale, E.; Superti-Furga, A.; Steinmann, B.; Ramirez, F.  
 J. Biol. Chem. 266, 5256-5259, 1991  
 A>Title: G to T transversion at position +5 of a splice donor site causes skipping of th  
 A;Reference number: I55349; MUID:91161621; PMID:1672129  
 A;Accession: I55349  
 A;Status: translated from GB/EMBL/DBJ  
 A;Molecule type: DNA  
 A;Residues: 537-605 <LEE>  
 A;Cross-references: GB:M59312; NID:g180815; PIDN:AA52041.1; PID:g180816  
 R;Seyer, J.M.; Mainardi, C.; Kang, A.H.  
 Biochemistry 19, 1583-1589, 1980  
 A>Title: Covalent structure of collagen: amino acid sequence of alpha1 (III)-CB5 from ty  
 A;Reference number: A90438; MUID:80198282; PMID:6246925  
 A;Accession: A90438  
 A;Molecule type: protein  
 A;Residues: 728-895, 'A', 897-964 <SEY4>  
 A;Experimental source: liver  
 R;Cole, W.G.; Chiodo, A.A.; Lamande, S.R.; Janeczko, R.; Ramirez, F.; Dahl, H.H.M.; Chan  
 J. Biol. Chem. 265, 17070-17077, 1990  
 A>Title: A base substitution at a splice site in the COL3A1 gene causes exon skipping an  
 A;Reference number: A38303; MUID:91009133; PMID:2145268  
 A;Accession: A38303  
 A;Molecule type: mRNA  
 A;Residues: 861-1015 <COL>  
 A;Cross-references: GB:J05617; GB:M55603; GB:M59227; NID:g180878; PIDN:AAB59383.1; PID:9  
 A;Note: a mutant sequence with 942-977 spliced out from a patient with Ehlers-Danlos syn  
 R;Mankoo, B.S.; Dalglish, R.  
 Nucleic Acids Res. 16, 2337, 1988  
 A>Title: Human pro alpha1(III) collagen: cDNA sequence for the 3' end.  
 A;Reference number: S02119; MUID:88189827; PMID:3357782  
 A;Accession: S02119

A;Status: translation not shown  
 A;Molecule type: mRNA  
 A;Residues: 950-1018, 'Y', 1020-1183, 'S', 1185-1466 <MAN>  
 A;Cross-references: EMBL:X06700; NID:g30053; PIDN:CAA29886.1; PID:g30054  
 R;Seyer, J.M.; Kang, A.H.  
 Biochemistry 20, 2621-2627, 1981  
 A>Title: Covalent structure of collagen: amino acid sequence of alpha1 (III)-CB9 from tyf  
 A;Reference number: A90446; MUID:81208139; PMID:7016180  
 A;Accession: A90446  
 A;Molecule type: protein  
 A;Residues: 965-979, 'A', 981-984, 'PS', 987, 'QN', 990-1096, 'P', 1098-1152, 'AT', 1155, 'S', 1157-1  
 A;Experimental source: liver  
 R;Loidl, H.R.; Brinker, J.M.; May, M.; Pihlajaniemi, T.; Morrow, S.; Rosenbloom, J.; Myer  
 Nucleic Acids Res. 12, 9383-9394, 1984  
 A>Title: Molecular cloning and carboxyl-propeptide analysis of human type III procollagen  
 A;Reference number: A93551; MUID:85087944; PMID:6096827  
 A;Accession: A93551  
 A;Molecule type: mRNA  
 A;Residues: 1065-1155, 'P', 1157-1466 <LOI>  
 A;Cross-references: EMBL:X01655; EMBL:X01742; NID:g29584; PIDN:CAA25821.1  
 R;Miskulin, M.; Dalglish, R.; Kluge-Becker, B.; Rennard, S.I.; Tolstoshev, P.; Brant  
 Biochemistry 25, 1408-1413, 1986  
 A>Title: Human type III collagen gene expression is coordinately modulated with the type  
 A;Reference number: I52393; MUID:86187804; PMID:3754462  
 A;Accession: I52393  
 A;Status: translated from GB/EMBL/DBJ  
 A;Molecule type: mRNA  
 A;Residues: 1161-1200 <MS>  
 A;Cross-references: GB:M13146; NID:g180415; PIDN:AA52003.1; PID:g180416  
 R;Emanuel, B.S.; Cannizzaro, L.A.; Seyer, J.M.; Myers, J.C.  
 Proc. Natl. Acad. Sci. U.S.A. 82, 3385-3389, 1985  
 A>Title: Human alpha 1(III) and alpha 2(V) procollagen genes are located on the long arm  
 A;Reference number: I59025; MUID:85216505; PMID:3858826  
 A;Accession: I59025  
 A;Status: translated from GB/EMBL/DBJ  
 A;Molecule type: mRNA  
 A;Residues: 1165-1196 <SWA>  
 A;Cross-references: GB:M11134; NID:g180417; PIDN:AA52004.1; PID:g180418  
 R;Chu, M.L.; Weil, D.; de Wet, W.; Bernard, M.; Sippola, M.; Ramirez, F.  
 J. Biol. Chem. 260, 4357-4363, 1985  
 A>Title: Isolation of cDNA and genomic clones encoding human pro-alpha(III) collagen. Pa  
 A;Reference number: A92516; MUID:85157600; PMID:2579949  
 A;Accession: A92516  
 A;Molecule type: DNA  
 A;Residues: 1176-1240, 'V', 1242-1356, 'P', 1358-1466 <CHU>  
 A;Cross-references: GB:M10615; GB:M10793; GB:M10794; GB:M10795; GB:M10796; GB:M10797; GB  
 A;Experimental source: liver  
 A;Note: the authors translated the codon TTC for residue 1057 as Tyr; the codons given fo  
 ation  
 C;Comment: Prolines and lysines at the third position of the tripeptide repeating unit (C  
 3-hydroxylated. About 15% of the lysines are 5-hydroxylated and some are subsequently O  
 C;Genetics:  
 A;Gene: GDB:COL3A1  
 A;Cross-references: GDB:118729; OMIM:120180  
 A;Map position: 2q31-2q31  
 A;Introns: 27/1; 94/3; 111/3; 149/3; 176/3; 554/3; 587/3; 1175/3; 1275/1; 1337/3; 1418/3  
 A;Note: the list of introns is incomplete; defects in this gene can result in Ehlers-Dan  
 C;Complex: type III collagen is a homotrimer of monomers initially linked by disulfide br  
 er of their length, is formed with desmosine cross-links made from lysine and allysine re  
 C;Function:  
 A;Description: structural component of extracellular fibrous polymer that maintains integ  
 C;Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;  
 C;Keywords: coiled coil; Ehlers-Danlos syndrome; extracellular matrix; glycoprotein; hyd  
 F;1-23/Domain: signal sequence #status predicted <SIG>  
 F;24-153/Domain: amino-terminal propeptide #status predicted <PRO>  
 F;31-91/Domain: von Willebrand factor type C repeat homology <WVC>  
 F;154-1221/Product: collagen alpha 1(III) chain #status predicted <WAT>  
 F;154-167/Region: amino-terminal nonhelical telopeptide  
 F;168-1196/Region: helical  
 F;1091-1093/Region: cell attachment (R-G-D) motif  
 F;1197-1221/Region: carboxyl-terminal nonhelical telopeptide  
 F;1222-1466/Domain: carboxyl-terminal propeptide #status predicted <CPR>  
 F;1238-1466/Domain: fibrillar collagen carboxyl-terminal homology <FCC>



F:24/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status predicted  
 F:153-154/Cleavage site: Pro-Gln (procollagen N-endopeptidase) #status predicted  
 F:154/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status predicted  
 F:161/1212/Modified site: allylsine (Lys) #status predicted  
 F:263/284,860,977,1106/Modified site: 5-hydroxyllysine (Lys) #status experimental  
 F:263/Binding site: carboxylate (Lys) (covalent) #status experimental  
 F:584,1094/Modified site: 5-hydroxyllysine (Lys) (partial) #status experimental  
 F:948-949/Cleavage site: Gly-Ile (collagenase) #status experimental  
 F:1106/Binding site: carbohydrate (Lys) (covalent) #status predicted

Query Match 100.0%; Score 42; DB 1; Length 1466;  
 Best Local Similarity 100.0%; Pred. No. 11;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IGGKAGG 8

DB 1208 IGGKAGG 1215

RESULT 2

S41067  
 collagen alpha 1(III) chain - rat  
 C:Species: Rattus norvegicus (Norway rat)  
 C:Date: 13-Jan-1995 #sequence revision 13-Jan-1995 #text\_change 13-Aug-1999  
 C:Accession: S41067; A29905; S31924  
 R:Glumoff, V.; Maekelae, J.K.; Vuorio, E.  
 Biochim. Biophys. Acta 1217, 41-48, 1994  
 A:Title: Cloning of cDNA for rat pro alpha-1(III) collagen mRNA. Different expression pa  
 A:Reference number: S41067; MUID:94114571; PMID:8286415  
 A:Accession: S41067  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-636 <GLU>  
 A:Cross-references: EMBL:X70369; NID:G57915; PIDN:CAA49832.1; PID:G57916  
 R:Frankel, F.R.; Heu, C.Y.J.; Myers, J.C.; Lin, E.; Lyttle, C.R.; Komm, B.; Mohn, K.  
 DNA 7, 347-354, 1988  
 A:Title: Regulation of alpha-2 (I), alpha-1 (III), and alpha-2 (V) collagen mRNAs by est  
 A:Reference number: A29905; MUID:88296083; PMID:2456904  
 A:Accession: A29905  
 A:Molecule type: mRNA  
 A:Residues: 308-482 <FRA>  
 A:Cross-references: GB:M21354; NID:G203500; PIDN:AAA40942.1; PID:G203501  
 R:Glumoff, V.; Maekelae, J.K.; Vuorio, E.  
 submitted to the EMBL Data Library, February 1993  
 A:Reference number: S31924  
 A:Accession: S31924  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 2-636 <GL2>  
 A:Cross-references: EMBL:X70369  
 C:Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;  
 C:Keywords: coiled coil; extracellular matrix; glycoprotein; trimer; triple helix  
 F:408-636/Domain: fibrillar collagen carboxyl-terminal homology <FCC>

Query Match 90.5%; Score 38; DB 2; Length 636;  
 Best Local Similarity 75.0%; Pred. No. 29;  
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 IGGKAGG 8

DB 378 VGGKSGG 385

RESULT 3

S5986  
 collagen alpha 1(III) chain precursor - mouse  
 C:Species: Mus musculus (house mouse)  
 C:Date: 10-Apr-1996 #sequence revision 19-Apr-1996 #text\_change 13-Aug-1999  
 C:Accession: S59856; S62120; S16373  
 R:Toman, P.D.; de Crombrughe, B.  
 Gene 147, 161-168, 1994  
 A:Title: The mouse type-III procollagen-encoding gene: genomic cloning and complete DNA  
 A:Reference number: S59856; MUID:95011609; PMID:7926795

A:Accession: S59856  
 A:Molecule type: DNA  
 A:Residues: 1-1464 <TOM>  
 A:Cross-references: EMBL:X52046  
 R:Toman, D.  
 submitted to the EMBL Data Library, November 1994  
 A:Reference number: S62120  
 A:Accession: S62120  
 A:Molecule type: DNA  
 A:Residues: 1-866, 'G', 868-1464 <TOA>

A:Cross-references: EMBL:X52046; NID:G575321; PIDN:CAA36279.1; PID:G575322  
 R:Metzger, M.; Toman, D.; de Crombrughe, B.; Vuorio, E.  
 Biochim. Biophys. Acta 1089, 241-243, 1991  
 A:Title: Specific hybridization probes for mouse type I, II, III and IX collagen mRNAs.  
 A:Reference number: S16176; MUID:91274355; PMID:2054384  
 A:Accession: S16373  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1442-1464 <MET>

A:Cross-references: EMBL:X57983; NID:G50476; PIDN:CAA41048.1; PID:G50477  
 C:Genetics:  
 A:Introns: 29/1; 95/3; 112/3; 150/3; 175/3; 193/3; 211/3; 229/3; 247/3; 265/3; 283/3; 298  
 58/3; 673/3; 706/3; 742/3; 760/3; 778/3; 796/3; 814/3; 850/3; 868/3; 886/3; 940/3; 976/3;  
 C:Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;  
 C:Keywords: coiled coil; extracellular matrix  
 F:1-24/Domain: signal sequence #status predicted <SIG>  
 F:25-154/Domain: propeptide #status predicted <PRO>  
 F:32-92/Domain: von Willibrand factor type C repeat homology <WVC>  
 F:155-1464/Product: collagen alpha 1(III) chain #status predicted <MAT>  
 F:1236-1464/Domain: fibrillar collagen carboxyl-terminal homology <FCC>

Query Match 90.5%; Score 38; DB 2; Length 1464;  
 Best Local Similarity 75.0%; Pred. No. 63;  
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 IGGKAGG 8

DB 1206 VGGKSGG 1213

RESULT 4

AG1031  
 hypothetical protein STY4577 [imported] - Salmonella enterica subsp. enterica serovar Ty  
 C:Species: Salmonella enterica subsp. enterica serovar Typhi  
 A:Note: this species has also been called Salmonella typhi  
 C:Date: 09-Nov-2001 #sequence\_revision 09-Nov-2001 #text\_change 18-Nov-2002  
 C:Accession: AG1031  
 R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,  
 th, T.; Connor, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,  
 S.; Moule, S.; O'Gaora, P.  
 Nature 413, 848-852, 2001  
 A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;  
 A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov  
 A:Reference number: AB0502; MUID:21534947; PMID:11677608  
 A:Accession: AG1031  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-476 <PAR>  
 A:Cross-references: GB:AL513382; PIDN:CAD09352.1; PID:G16505352; GSPDB:GN00176  
 C:Genetics:  
 A:Gene: STY4577

Query Match 88.1%; Score 37; DB 2; Length 476;  
 Best Local Similarity 75.0%; Pred. No. 34;  
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 IGGKAGG 8

DB 211 VGGKAGG 218

RESULT 5

S56535

hypothetical 43.7K protein (fecI-fimB intergenic region) - Escherichia coli (strain K-12)  
 N/Alternate names: hypothetical protein f404  
 C/Species: Escherichia coli  
 C/Date: 28-Oct-1995 #sequence\_revision 03-Nov-1995 #text\_change 01-Mar-2002  
 C/Accession: S56535; H65244  
 R/Burland, V.; Plunkett III, G.; Sofia, H.J.; Daniels, D.L.; Blattner, F.R.  
 Nucleic Acids Res. 23, 2105-2119, 1995  
 A/Title: Analysis of the Escherichia coli genome VI: DNA sequence of the region from 92.  
 A/Reference number: S56314; MUID:95334362; PMID:7610040  
 A/Accession: S56535  
 A/Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A/Molecule type: DNA  
 A/Residues: 1-404 <BUR>  
 A/Cross-references: EMBL:U14003; NID:gl263172; PIDN:AAA97206.1; PID:9537151  
 A/Note: the nucleotide sequence was submitted to the EMBL Data Library, August 1994  
 R/Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co.  
 A.; Rose, D.J.; Mau, B.; Shao, Y.  
 Science 277, 1453-1462, 1997  
 A/Title: The complete genome sequence of Escherichia coli K-12.  
 A/Reference number: A64720; MUID:97426617; PMID:9278503  
 A/Accession: H65244  
 A/Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A/Molecule type: DNA  
 A/Residues: 1-404 <BLAT>  
 A/Cross-references: GB:AE000501; GB:U00096; NID:g23637372; PIDN:BAC77266.1; PID:gl790764;  
 A/Experimental source: strain K-12, substrain MG1655  
 C/Genetics:  
 A/Gene: yjht

Query Match 85.7%; Score 36; DB 2; Length 404;  
 Best Local Similarity 87.5%; Pred. No. 44;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 Qy 1 IGGEKAGG 8  
 |||||  
 Db 377 IGGETAGG 384

RESULT 6  
 E91287  
 hypothetical protein ECs5269 [imported] - Escherichia coli (strain O157:H7, substrain R1  
 C/Species: Escherichia coli  
 C/Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 18-Jul-2001  
 C/Accession: E91287  
 R/Hayashi, T.; Makino, K.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.  
 Gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.  
 DNA Res. 8, 11-22, 2001  
 A/Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and geno  
 A/Reference number: A99629; MUID:21156231; PMID:11258796  
 A/Accession: E91287  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-404 <HAY>  
 A/Cross-references: PIDN:BAB38692.1; PID:gl13364747; GSPDB:GN00154  
 A/Experimental source: strain O157:H7, substrain RMD 0509952  
 C/Genetics:  
 A/Gene: ECs5269

Query Match 85.7%; Score 36; DB 2; Length 404;  
 Best Local Similarity 87.5%; Pred. No. 44;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 Qy 1 IGGEKAGG 8  
 |||||  
 Db 377 IGGETAGG 384

RESULT 7  
 H86128  
 hypothetical protein yjht [imported] - Escherichia coli (strain O157:H7, substrain EDL93  
 C/Species: Escherichia coli  
 C/Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 14-Sep-2001  
 C/Accession: H86128

R/Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew,  
 iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamouisis, K.; Apodaca,  
 Nature 409, 529-533, 2001  
 A/Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.  
 A/Reference number: A85480; MUID:21074935; PMID:11206551  
 A/Accession: H86128  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-404 <STO>  
 A/Cross-references: GB:AB005174; NID:gl2519317; PIDN:AAG59492.1; GSPDB:GN00145; UWGP:Z59  
 A/Experimental source: strain O157:H7, substrain EDL933  
 C/Genetics:  
 A/Gene: yjht

Query Match 85.7%; Score 36; DB 2; Length 404;  
 Best Local Similarity 87.5%; Pred. No. 44;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 Qy 1 IGGEKAGG 8  
 |||||  
 Db 377 IGGETAGG 384

RESULT 8  
 S14998  
 heat shock protein 18 (clone c9) - maize  
 C/Species: Zea mays (maize)  
 C/Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 13-Aug-1999  
 C/Accession: S14998  
 R/Goping, I.S.; Prappier, J.R.H.; Walden, D.B.; Atkinson, B.G.  
 Plant Mol. Biol. 16, 699-711, 1991  
 A/Title: Sequence, identification and characterization of cDNAs encoding two different m  
 A/Reference number: S14997; MUID:91329703; PMID:1714322  
 A/Accession: S14998  
 A/Molecule type: mRNA  
 A/Residues: 1-161 <GOP>  
 A/Cross-references: EMBL:X54076; NID:g22338; PIDN:CAA38013.1; PID:g223339  
 C/Genetics:  
 A/Gene: hsp18-9  
 C/Superfamily: alpha-crystallin  
 C/Keywords: heat shock; stress-induced protein

Query Match 83.3%; Score 35; DB 2; Length 161;  
 Best Local Similarity 85.7%; Pred. No. 29;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 Qy 2 GGEXKAGG 8  
 |||||  
 Db 30 GGDKAGG 36

RESULT 9  
 H69074  
 formylmethanofuran dehydrogenase (EC 1.2.99.5) (tungsten) chain C - Methanobacterium the  
 C/Species: Methanobacterium thermoautotrophicum  
 C/Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 24-Sep-1999  
 C/Accession: H69074  
 R/Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.; Dubois, J.; Aldredge, T.; E  
 ; Qiu, D.; Spadafora, R.; Vicaire, R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiwani, N.  
 ki, S.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.  
 J. Bacteriol. 179, 7135-7155, 1997  
 A/Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta H: funct  
 A/Reference number: A69000; MUID:98037514; PMID:9371463  
 A/Accession: H69074  
 A/Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A/Molecule type: DNA  
 A/Residues: 1-270 <MTH>  
 A/Cross-references: GB:AB000916; GB:AB000666; NID:g2622674; PIDN:AAB86032.1; PID:g262268  
 A/Experimental source: strain Delta H  
 C/Genetics:  
 A/Gene: MTH1558  
 C/Superfamily: formylmethanofuran dehydrogenase chain C  
 C/Keywords: iron-sulfur protein; metalloprotein; oxidoreductase; tungsten

Query Match 85.7%; Score 36; DB 2; Length 404;  
 Best Local Similarity 87.5%; Pred. No. 44;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 Qy 1 IGGEKAGG 8  
 |||||  
 Db 377 IGGETAGG 384

Query Match 83.3%; Score 35; DB 2; Length 270;  
 Best Local Similarity 75.0%; Pred. No. 47;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 IGGEKAGG 8  
 :|||:|  
 Db 200 VGEWAGG 207

RESULT 10  
 S57457  
 formylmethanofuran dehydrogenase (EC 1.2.99.5) (tungsten) chain C - Methanobacterium the  
 C:Species: Methanobacterium thermoautotrophicum  
 C:Date: 10-Oct-1995 #sequence\_revision 17-Nov-1995 #text\_change 20-Jun-2000  
 C:Accession: S63546; S57457  
 R:Hochheimer, A.; Schmitz, R.A.; Thauer, R.K.; Hedderich, R.  
 Eur. J. Biochem. 234, 910-920, 1995  
 A:Title: The tungsten formylmethanofuran dehydrogenase from Methanobacterium thermoautot  
 A:Reference number: S63519; PMID:96163477; PMID:8575452  
 A:Accession: S63546  
 A:Molecule type: DNA  
 A:Residues: 1-270 <HOC>  
 A:Cross-references: EMBL:X87970; NID:G1890205; PIDN:CAA61214.1; PID:G1890211  
 A:Experimental source: strain Marburg, DSM 2133  
 C:Genetics:  
 A:Gene: fwdC  
 C:Complex: heterotetramer of A (see PIR:S57456), B (see PIR:S57458), C, and D (see PIR:S  
 C:Superfamily: formylmethanofuran dehydrogenase chain C  
 C:Keywords: heterotetramer; iron-sulfur protein; metalloprotein; oxidoreductase; tungste

Query Match 83.3%; Score 35; DB 2; Length 270;  
 Best Local Similarity 75.0%; Pred. No. 47;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 IGGEKAGG 8  
 :|||:|  
 Db 200 VGEWAGG 207

RESULT 11  
 AE2301  
 phosphoglucomutase/phosphomannomutase [imported] - Nostoc sp. (strain PCC 7120)  
 C:Species: Nostoc sp. PCC 7120  
 A:Nore: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120  
 C:Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 09-Dec-2002  
 C:Accession: AE2301  
 R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi  
 Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S  
 DNA Res. 8, 205-213, 2001  
 A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana  
 A:Reference number: AB1807; PMID:21595285; PMID:11759840  
 A:Accession: AE2301  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-475 <KUR>  
 A:Cross-references: GB:BA000019; PIDN:BA075663.1; PID:G17133098; GSPDB:GN00179  
 A:Experimental source: strain PCC 7120  
 C:Genetics:  
 C:Superfamily: phosphomannomutase

Query Match 83.3%; Score 35; DB 2; Length 475;  
 Best Local Similarity 75.0%; Pred. No. 79;  
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 IGGEKAGG 8  
 :|||:|  
 Db 332 IGGEESGG 339

RESULT 12  
 S75539

hypothetical protein slr1334 - Synechocystis sp. (strain PCC 6803)  
 C:Species: Synechocystis sp.

A:Variety: PCC 6803  
 C:Date: 25-Apr-1997 #sequence\_revision 25-Apr-1997 #text\_change 20-Jun-2000  
 C:Accession: S75539  
 R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;  
 O, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda,  
 DNA Res. 3, 109-136, 1996  
 A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis  
 s.

A:Reference number: S74322; PMID:97061201; PMID:8905231

A:Accession: S75539  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-490 <KAN>  
 A:Cross-references: EMBL:D90911; GB:AB001339; NID:G1653083; PIDN:BA018100.1; PID:G1653184  
 A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996  
 C:Superfamily: phosphomannomutase

Query Match 83.3%; Score 35; DB 2; Length 490;  
 Best Local Similarity 75.0%; Pred. No. 81;  
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 IGGEKAGG 8  
 :|||:|  
 Db 337 IGGEESGG 344

RESULT 13  
 T41103  
 60S ribosomal protein l12 - fission yeast (Schizosaccharomyces pombe)  
 C:Species: Schizosaccharomyces pombe  
 C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 31-Jan-2000  
 C:Accession: T41103; T41292  
 R:Purnelle, B.; Goffeau, A.; Wood, V.; Rajandream, M.A.; Barrell, B.G.; Rieger, M.  
 submitted to the EMBL Data Library, September 1998  
 A:Reference number: Z21964  
 A:Accession: T41103  
 A:Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: DNA  
 A:Residues: 1-165 <PUR>  
 A:Cross-references: EMBL:AL031535; PIDN:CAA20752.1; GSPDB:GN00068; SPDB:SPCC16C4.13C  
 A:Experimental source: strain 972h-; cosmid c16C4  
 R:Wood, V.; Rajandream, M.A.; Barrell, B.G.; Rieger, M.  
 submitted to the EMBL Data Library, September 1998  
 A:Reference number: Z21985  
 A:Accession: T41292

A:Status: preliminary; translated from GB/EMBL/DDBJ  
 A:Molecule type: DNA  
 A:Residues: 1-165 <WOO>  
 A:Cross-references: EMBL:AL031824; PIDN:CAA21221.1; GSPDB:GN00068; SPDB:SPCC31H12.04C  
 A:Experimental source: strain 972h-; cosmid c31H12  
 C:Genetics:  
 A:Gene: SPCC16C4.13c; SPCC31H12.04c  
 A:Map position: 3  
 A:Introns: 4/3  
 C:Superfamily: rat ribosomal protein L12

Query Match 81.0%; Score 34; DB 2; Length 165;  
 Best Local Similarity 75.0%; Pred. No. 45;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 IGGEKAGG 8  
 :|||:|  
 Db 18 VGEWAGG 25

RESULT 14  
 B36724  
 hypothetical protein 2 (glnII 5' region) - Streptomyces viridochromogenes  
 C:Species: Streptomyces viridochromogenes  
 C:Date: 12-Apr-1991 #sequence\_revision 12-Apr-1991 #text\_change 22-Oct-1999  
 C:Accession: B36724; S09194; S11634

R;Behrmann, I.; Hillemann, D.; Puehler, A.; Strauch, E.; Wohlleben, W.  
 J. Bacteriol. 172, 5326-5334, 1990  
 A;Title: Overexpression of a Streptomyces viridochromogenes gene (glnII) encoding a glut  
 yl-alanine.  
 A;Reference number: A36724; MUID:90368592; PMID:1975583  
 A;Accession: B36724  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-240 <BEH>  
 A;Cross-references: EMBL:X52842; NID:g47983; PIDN:CAA37027.1; PID:g47985

Query Match 81.0%; Score 34; DB 2; Length 240;  
 Best Local Similarity 85.7%; Pred.No. 64;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 2 GGEKAGG 8  
 |||:||||  
 Db 118 GGEQAGG 124

RESULT 15  
 T31264  
 cis-1,2-dihydro-1,2-dihydroxynaphthalene dehydrogenase homolog - Sphingomonas aromaticiv  
 C;Species: Sphingomonas aromaticivorans  
 C;Date: 11-Jan-2000 #sequence\_revision 11-Jan-2000 #text\_change 04-Mar-2000  
 C;Accession: T31264  
 R;Romine, M.F.; Stillwell, L.C.; Wong, K.K.; Thurston, S.J.; Sisk, E.C.; SENSEN, C.W.; G  
 submitted to the EMBL Data Library, July 1998  
 A;Description: Complete sequence of a 184 kb catabolic plasmid from Sphingomonas aromati  
 A;Reference number: Z20992  
 A;Accession: T31264  
 A;Status: preliminary; translated from GB/EMBL/DBJ  
 A;Molecule type: DNA  
 A;Residues: 1-266 <ROM>  
 A;Cross-references: EMBL:AF079317; NID:g3378261; PID:g3378405; PIDN:AAD03988.1  
 C;Genetics:  
 A;Genome: plasmid pNL1  
 A;Note: bpH  
 C;Superfamily: ribitol dehydrogenase; short-chain alcohol dehydrogenase homology

Query Match 81.0%; Score 34; DB 2; Length 266;  
 Best Local Similarity 75.0%; Pred.No. 70;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 IGGEKAGG 8  
 :|||:||||  
 Db 193 LGGTKAGG 200

Search completed: September 18, 2004, 04:30:50  
 Job time : 11.7255 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 18, 2004, 03:48:54 ; Search time 5.96078 Seconds  
(without alignments)  
69.884 Million cell updates/sec

Title: US-10-615-959-40  
Perfect score: 42  
Sequence: 1 IGGEKAGG 8

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_42.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	42	100.0	1466	1	P02461 homo sapien
2	38	90.5	636	1	P13941 rattus norv
3	38	90.5	1464	1	P08121 mus musculus
4	36	85.7	368	1	P39371 escherichia
5	35	83.3	161	1	P24631 zea mays (m
6	35	83.3	270	1	FWDG_METTH
7	35	83.3	270	1	FWDG_METTM
8	35	83.3	559	1	MERA_ALCSP
9	34	81.0	165	1	RL12_SCHPO
10	34	81.0	240	1	YGL2_STRVR
11	34	81.0	1137	1	RLR1_HSV11
12	33	78.6	312	1	YD63_THEAC
13	33	78.6	374	1	TRMU_VIBPA
14	33	78.6	374	1	TRMU_VIBVU
15	33	78.6	1145	1	DP2L_METAC
16	32	76.2	155	1	DHB_GAPL
17	32	76.2	270	1	FWDG_METWO
18	32	76.2	275	1	YX21_CAUCR
19	32	76.2	287	1	RL6_HUMAN
20	32	76.2	287	1	RL6_MOUSE
21	32	76.2	296	1	RL6_RAT
22	32	76.2	302	1	HSJO_NEIVA
23	32	76.2	302	1	HSJO_NEIBW
24	32	76.2	384	1	YH74_VIBCH
25	32	76.2	528	1	C318_DROME
26	32	76.2	917	1	FTSK_STRCO
27	32	76.2	1013	1	Y793_DEIRA
28	32	76.2	1332	1	XKDO_BACSU
29	32	76.2	1522	1	ARHB_HUMAN
30	32	76.2	1527	1	ARHB_RAT
31	31	73.8	145	1	URI_CYPCA
32	31	73.8	188	1	RL13_LUMRU
33	31	73.8	212	1	RB2A_HUMAN

34	31	73.8	212	1	RB2A_MOUSE
35	31	73.8	212	1	RB2A_RABIT
36	31	73.8	212	1	RB2A_RAT
37	31	73.8	248	1	CYPB_VICFA
38	31	73.8	371	1	TRMU_VERPE
39	31	73.8	486	1	LIP1_YARLI
40	31	73.8	553	1	DMWD_HUMAN
41	31	73.8	650	1	DMWD_MOUSE
42	31	73.8	1679	1	FUR2_DROME
43	31	73.8	2485	1	PTND_HUMAN
44	31	73.8	3530	1	MYL5_HUMAN
45	30	71.4	49	1	YKQ6_CAEEL

## ALIGNMENTS

RESULT 1  
CAL3\_HUMAN  
ID CAL3\_HUMAN STANDARD; PRT; 1466 AA.  
AC P02461; Q15112;  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 01-JAN-1990 (Rel. 13, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Collagen alpha 1(III) chain precursor.  
GN COL3A1.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Skin fibroblast;  
RX MEDLINE=89350838; PubMed=2764886;  
RA Ala-Kokko L., Kontusaari S., Baldwin C.T., Kuivaniemi H.,  
RA Prockop D.J.;  
RT "Structure of cDNA clones coding for the entire prepro alpha 1 (III)  
chain of human type III procollagen. Differences in protein structure  
from type I procollagen and conservation of codon preferences.";  
RL Biochem. J. 260:509-516(1989).  
RN [2]  
RP SEQUENCE OF 149-1225 FROM N.A.  
RX MEDLINE=89386015; PubMed=2780304;  
RA Janeczko R.A., Ramirez F.;  
RT "Nucleotide and amino acid sequences of the entire human alpha 1  
(III) collagen.";  
RL Nucleic Acids Res. 17:6742-6742(1989).  
RN [3]  
RP SEQUENCE OF 168-398.  
RX MEDLINE=77134724; PubMed=557335;  
RA Seyer J.M., Kang A.H.;  
RT "Covalent structure of collagen: amino acid sequence of cyanogen  
bromide peptides from the amino-terminal segment of type III collagen  
of human liver.";  
RL Biochemistry 16:1158-1164(1977).  
RN [4]  
RP REVISIONS.  
RA Seyer J.M.;  
RL Submitted (DEC-1977) to the PIR data bank.  
RN [5]  
RP SEQUENCE OF 399-727.  
RX MEDLINE=79000343; PubMed=687591;  
RA Seyer J.M., Kang A.H.;  
RT "Covalent structure of collagen: amino acid sequence of five  
consecutive CNBr peptides from type III collagen of human liver.";  
RL Biochemistry 17:3404-3411(1978).  
RN [6]  
RP SEQUENCE OF 728-964.  
RX MEDLINE=80198282; PubMed=6246925;  
RA Seyer J.M., Mainardi C., Kang A.H.;  
RT "Covalent structure of collagen: amino acid sequence of alpha 1  
(III)-CB5 from type III collagen of human liver.";  
RL Biochemistry 19:1583-1589(1980).

[7] RN SEQUENCE OF 950-1466 FROM N.A.  
 RP MEDLINE=88189827; PubMed=3357782;  
 RA Mankoo B.S., Dalgleish R.;  
 RT "Human pro alpha 1(III) collagen: cDNA sequence for the 3' end.";  
 RL Nucleic Acids Res. 16:2337-2337(1988).  
 [8] RN REVISION TO 1184.  
 RP MEDLINE=89098346; PubMed=3211760;  
 RA Molyneux K., Dalgleish R.;  
 RT "Human type III collagen 'variant' is a cDNA cloning artefact.";  
 RL Nucleic Acids Res. 16:1183-1183(1988).  
 [9] RN SEQUENCE OF 1065-1466 FROM N.A.  
 RP MEDLINE=85087944; PubMed=6096827;  
 RA Loidi H.R., Brinker J.M., May M., Pihlajaniemi T., Morrow S.,  
 RA Rosenbloom J., Myers J.C.;  
 RT "Molecular cloning and carboxyl-propeptide analysis of human type III procollagen.";  
 RL Nucleic Acids Res. 12:9383-9394(1984).  
 [10] RN SEQUENCE OF 965-1200.  
 RP MEDLINE=81208139; PubMed=7016180;  
 RA Seyer J.M., Kang A.H.;  
 RT "Covalent structure of collagen: amino acid sequence of alpha 1(IIII)-C89 from type III collagen of human liver.";  
 RL Biochemistry 20:2621-2627(1981).  
 [11] RN SEQUENCE OF 1176-1466 FROM N.A.  
 RP MEDLINE=85157600; PubMed=2579949;  
 RA Chu M.-L., Weil D., de Wet W.J., Bernard M.P., Sippola M., Ramirez F.;  
 RT "Isolation of cDNA and genomic clones encoding human pro-alpha 1 gene.";  
 RL J. Biol. Chem. 260:4357-4363(1985).  
 [12] RN SEQUENCE OF 1161-1200 FROM N.A.  
 RP MEDLINE=86187804; PubMed=3754462;  
 RA Miskulin M., Dalgleish R., Klueve-Beckerman B., Rennard S.I.,  
 RA Tolstoshev P., Brantly M., Crystal R.G.;  
 RT "Human type III collagen gene expression is coordinately modulated with the type I collagen genes during fibroblast growth.";  
 RL Biochemistry 25:1408-1413(1986).  
 [13] RN SEQUENCE OF 1-170 FROM N.A.  
 RP TISSUE=Placenta;  
 RP MEDLINE=88303360; PubMed=3405773;  
 RA Tomar D., Ricca G., de Crombrughe B.;  
 RT "Nucleotide sequence of a cDNA coding for the amino-terminal region of human prepro alpha 1(III) collagen.";  
 RL Nucleic Acids Res. 16:7201-7201(1988).  
 [14] RN SEQUENCE OF 1-176 FROM N.A.  
 RP MEDLINE=89378752; PubMed=2777083;  
 RA Benson-Chanda V., Su M.W., Weil D., Chu M.-L., Ramirez F.;  
 RT "Cloning and analysis of the 5' portion of the human type-III procollagen gene (COL3A1).";  
 RL Gene 78:255-265(1989).  
 [15] RN REVIEW ON VARIANTS.  
 RP MEDLINE=97255959; PubMed=9101290;  
 RA Kuivaniemi H., Tromp G., Prockop D.J.;  
 RT "Mutations in fibrillar collagens (Types I, II, III, and XI), fibrillar-associated collagen (type IX), and network-forming collagen (type X) cause a spectrum of diseases of bone, cartilage, and blood vessels.";  
 RL Hum. Mutat. 9:300-315(1997).  
 [16] RN VARIANT AORTIC ANEURYSM ARG-303, AND VARIANT THR-668.  
 RP MEDLINE=93293988; PubMed=8514866;  
 RA Tromp G., Wu Y., Prockop D.J., Madhathari S.L., Kleiner C.,  
 RA Earley J.J., Zhuang J., Noerregaard O., Darling R.C., Abbott W.M.,  
 RA Cole C.W., Jaakkola P., Rynanen M., Pearce W.H., Yao J.S.T.,  
 RA Majamaa K., Smuileens S.N., Gatalica Z., Ferrell R.E., Jimenez S.A.,  
 RA Jackson C.E., Michels V.V., Kaye M., Kuivaniemi H.;  
 RT "Sequencing of cDNA from 50 unrelated patients reveals that mutations in the triple-helical domain of type III procollagen are an infrequent cause of aortic aneurysms.";  
 RL J. Clin. Invest. 91:2539-2545(1993).  
 [17] RN VARIANT THR-698.  
 RP MEDLINE=91045136; PubMed=2235526;  
 RA Zafarullah K., Kleiner C., Tromp G., Kuivaniemi H., Kontusaari S.,  
 RA Wu Y., Ganguly A., Prockop D.J.;  
 RT "G to A polymorphism in exon 31 of the COL3A1 gene.";  
 RL Nucleic Acids Res. 18:6180-6180(1990).  
 [18] RN VARIANT AORTIC ANEURYSM ARG-786.  
 RP MEDLINE=91056145; PubMed=2243125;  
 RA Kontusaari S., Tromp G., Kuivaniemi H., Romanic A.M., Prockop D.J.;  
 RT "A mutation in the gene for type III procollagen (COL3A1) in a family with aortic aneurysms.";  
 RL J. Clin. Invest. 86:1465-1473(1990).  
 [19] RN VARIANT EDS-IV ARG-828.  
 RP MEDLINE=94016385; PubMed=8411057;  
 RA Richards A.J., Narcisi P., Lloyd J.C., Ferguson C., Pope F.M.;  
 RT "The substitution of glycine 661 by arginine in type III collagen produces mutant molecules with different thermal stabilities and causes Ehlers-Danlos syndrome type IV.";  
 RL J. Med. Genet. 30:690-693(1993).  
 [20] RN VARIANT EDS-IV SER-957.  
 RP MEDLINE=89109135; PubMed=2492273;  
 RA Tromp G., Kuivaniemi H., Shikata H., Prockop D.J.;  
 RT "A single base mutation that substitutes serine for glycine 790 of the alpha 1(III) chain of type III procollagen exposes an arginine and causes Ehlers-Danlos syndrome IV.";  
 RL J. Biol. Chem. 264:1349-1352(1989).  
 [21] RN VARIANT EDS-IV VAL-960.  
 RP MEDLINE=95268429; PubMed=7749417;  
 RA Tromp G., de Paape A., Nuytink L., Madhathari S.L., Kuivaniemi H.;  
 RT "Substitution of valine for glycine 793 in type III procollagen in Ehlers-Danlos syndrome type IV.";  
 RL Hum. Mutat. 5:179-181(1995).  
 [22] RN VARIANT EDS-IV GLU-1014.  
 RP MEDLINE=92316511; PubMed=1352273;  
 RA Richards A.J., Ward P.N., Narcisi P., Nicholls A.C., Lloyd J.C., Pope F.M.;  
 RT "A single base mutation in the gene for type III collagen (COL3A1) converts glycine 847 to glutamic acid in a family with Ehlers-Danlos syndrome type IV. An unaffected family member is mosaic for the mutation.";  
 RL Hum. Genet. 89:414-418(1992).  
 [23] RN VARIANT EDS-IV ASP-1050.  
 RP MEDLINE=90037070; PubMed=2808425;  
 RA Tromp G., Kuivaniemi H., Stolle C.A., Pope F.M., Prockop D.J.;  
 RT "Single base mutation in the type III procollagen gene that converts the codon for glycine 883 to aspartate in a mild variant of Ehlers-Danlos syndrome IV.";  
 RL J. Biol. Chem. 264:19313-19317(1989).  
 [24] RN VARIANT EDS-IV VAL-1077.  
 RP MEDLINE=91374480; PubMed=1895316;  
 RA Richards A.J., Lloyd J.C., Ward P.N., de Paape A., Narcisi P., Pope F.M.;  
 RT "Characterisation of a glycine to valine substitution at amino acid position 910 of the triple helical region of type III collagen in a patient with Ehlers-Danlos syndrome type IV.";  
 RL J. Med. Genet. 28:458-463(1991).  
 [25] RN VARIANT EDS-IV GLU-1173.  
 RP MEDLINE=93022543; PubMed=1357232;  
 RA Johnson P.H., Richards A.J., Pope F.M., Hopkinson D.A.;

DR SMART: SMO00038; COLFI: 1.  
DR PROSITE; PS01208; WAFc\_1; PARTIAL.  
KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;  
KW Collagen; Glycoprotein.  
FT NON TER 1 1  
FT CHAIN <1 375 COLLAGEN ALPHA 1(III) CHAIN.  
FT PROPEP 376 636 CARBOXYL-TERMINAL PROPEPTIDE.  
FT DOMAIN <1 368 TRIPLE-HELICAL REGION.  
FT DOMAIN 369 636 NONHELICAL REGION (C-TERMINAL).  
FT DISULFID 368 368 INTERCHAIN (BY SIMILARITY).  
FT DISULFID 369 369 INTERCHAIN (BY SIMILARITY).  
FT CONFLICT 340 340 N -> D (IN REF. 2).  
FT CONFLICT 429 429 A -> G (IN REF. 2).  
SQ SEQUENCE 636 AA; 62332 MW; 61A48159F01D01EE CRC64;

Query Match 90.5%; Score 38; DB 1; Length 636;  
Best Local Similarity 75.0%; Pred. No. 14;  
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps

QY 1 IGGEKAGG 8  
:|||||  
Db 378 VGGEKSGG 385

RESULT 3

Ca13 MOUSE STANDARD; PRT; 1464 AA.

ID CA13 MOUSE Q9CRN7;  
AC P08121; Q61429; Q9CRN7;  
DT 01-AUG-1988 (Rel. 08, Created)  
DT 15-JUN-1999 (Rel. 38, Last sequence update)  
DT 15-MAR-2004 (Rel. 43, Last annotation update)  
DE Collagen alpha 1(III) chain precursor.  
GN COL3A1.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6 X DBA; TISSUE=Embryo;  
RX MEDLINE=95011609; PubMed=7926795;  
RA Toman D., de Crombrughe B.;  
RT "The mouse type-III procollagen-encoding gene: genomic cloning and  
RT complete DNA sequence.";  
RL Gene 147:161-168(1994).  
RP [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6; TISSUE=Brain;  
RX MEDLINE=22388257; PubMed=12477932;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
RA Rana S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,  
RA Bosak R.A., McSwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahey J., Helton E., Ketterman M., Madan A., Rodrigues S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakesley R., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalls D.E.,  
RA Schnurch A., Schein J.E., Jones S.J.M., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length  
RT human and mouse cDNA sequences";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
RN [3]  
RP SEQUENCE OF 1-488 FROM N.A.  
RX MEDLINE=88167858; PubMed=3443309;  
RA Wood L., Theriault N., Voqeli G.;

RT "Complete nucleotide sequence of the N-terminal domains of the murine  
 RT alpha-1 type-III collagen chain.";  
 RL Gene 61:225-230(1987).  
 RN [4]

RP SEQUENCE OF 1-28 FROM N.A.  
 RX MEDLINE=85131189; PubMed=3972847;  
 RA Liau G., Mudryj M., de Crombrughe B.;  
 RT "Identification of the promoter and first exon of the mouse alpha 1  
 RT (III) collagen gene.";  
 RL J. Biol. Chem. 260:3773-3777 (1985).  
 RN [5]

RP SEQUENCE OF 810-1464 FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Embryonic head;  
 RX MEDLINE=21085660; PubMed=11217851;  
 RA Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
 RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,  
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,  
 RA Saito T., Okazaki Y., Gotohori T., Bono H., Kasukawa T., Saito R.,  
 RA Kadoya K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,  
 RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,  
 RA Schriml L.M., Staehli F., Suzuki R., Tomita M., Wagner L., Washio T.,  
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,  
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,  
 RA Brownstein M.J., Dutt C., Fletcher C., Fujita M., Gariboldi M.,  
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,  
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,  
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,  
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,  
 RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,  
 RA Hayashizaki Y.;  
 RT "Functional annotation of a full-length mouse cDNA collection.";  
 RL Nature 409:685-690(2001).  
 RN [6]

RP SEQUENCE OF 1442-1464 FROM N.A.  
 RC STRAIN=C57BL/6J;  
 RX MEDLINE=91274355; PubMed=2054384;  
 RA Metsaeranta M., Toman D., de Crombrughe B., Vuorio E.;  
 RT "Specific hybridization probes for mouse type I, II, III and IX  
 RT collagen mRNAs.";  
 RL Biochim. Biophys. Acta 1089:241-243(1991).  
 CC [-] FUNCTION: Collagen type III occurs in most soft connective tissues

CC along with type I collagen.

CC [-] SUBUNIT: Trimers of identical alpha 1(III) chains. The chains are  
 CC linked to each other by interchain disulfide bonds. Trimers are  
 CC also cross-linked via hydroxylsines.

CC [-] PTM: Proline residues at the third position of the tripeptide  
 CC repeating unit (G-X-Y) are hydroxylated in some or all of the  
 CC chains.

CC [-] PTM: O-linked glycan consists of a Glc-Gal disaccharide bound to  
 CC the oxygen atom of a post-translationally added hydroxyl group (By  
 CC similarity).

CC [-] SIMILARITY: Contains 1 VWFC domain.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

CC EMBL; X52046; CAA36279.1; -

CC EMBL; BC043089; AAH43089.1; -

CC EMBL; BC058724; AAH58724.1; -

CC EMBL; M18933; AAA37338.1; -

CC EMBL; AK03037; -; NOT ANNOTATED - CDS.

CC EMBL; AK019448; BAB31724.1; -

CC EMBL; X57983; CAA41048.1; -

CC PIR; A27353; A27353.

CC PIR; S59856; S59856.

CC MGD; MGI:88453; Col3a1.

DR InterPro; IPR008161; Clg\_helix.  
 DR InterPro; IPR008180; Collagen.  
 DR InterPro; IPR008885; Fib\_collagen\_C.  
 DR InterPro; IPR002181; Fibrinogen\_C.  
 DR InterPro; IPR001007; VWFC\_C.  
 DR Pfam; PF01410; COLFI; 1.  
 DR Pfam; PF01391; Collagen; 18.  
 DR ProDom; PD000007; Clg\_helix; 1.  
 DR ProDom; PD002078; Fib\_collagen\_C; 1.  
 DR SMART; SM00038; COLFI; 1.  
 DR SMART; SM00214; VWFC; 1.  
 DR PROSITE; PS01208; VWFC 1; 1.  
 DR PROSITE; PS0184; VWFC 2; 1.  
 KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;  
 KW Glycoprotein; Collagen; Signal.  
 FT SIGNAL 1 23 BY SIMILARITY.  
 FT PROPEP 24 154 AMINO-TERMINAL PROPEPTIDE.  
 FT CHAIN 155 1203 COLLAGEN ALPHA 1(III) CHAIN.  
 FT PROPEP 1204 1464 CARBOXYL-TERMINAL PROPEPTIDE.  
 FT DOMAIN 31 90 VWFC.  
 FT DOMAIN 155 169 NONHELICAL REGION (N-TERMINAL).  
 FT DOMAIN 170 1195 TRIPLE-HELICAL REGION.  
 FT DOMAIN 1196 1464 NONHELICAL REGION (C-TERMINAL).  
 FT CARBOHYD 262 262 O-LINKED (GAL. .) (BY SIMILARITY).  
 FT MOD\_RES 262 262 HYDROXYLATION (BY SIMILARITY).  
 FT MOD\_RES 283 283 HYDROXYLATION (BY SIMILARITY).  
 FT MOD\_RES 859 859 HYDROXYLATION (BY SIMILARITY).  
 FT MOD\_RES 976 976 HYDROXYLATION (BY SIMILARITY).  
 FT MOD\_RES 1093 1093 HYDROXYLATION (BY SIMILARITY).  
 FT MOD\_RES 1105 1105 HYDROXYLATION (BY SIMILARITY).  
 FT DISULFID 1195 1195 INTERCHAIN (BY SIMILARITY).  
 FT DISULFID 1196 1196 INTERCHAIN (BY SIMILARITY).  
 SQ SEQUENCE 1464 AA; 138944 MW; 2104EC27A886090B CRC64;

Query Match 90.5%; Score 38; DB 1; Length 1464;  
 Best Local Similarity 75.0%; Pred. No. 31;  
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 IGCEKAGG 8

Db 1206 VGGEKSGG 1213

RESULT 4

YJHT\_ECOLI STANDARD; PRT; 368 AA.  
 ID YJHT\_ECOLI  
 AC P39371;  
 DT 01-FEB-1995 (Rel. 31, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Hypothetical protein yJHT precursor.  
 GN YJHT OR B4310.  
 OS Escherichia coli.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Escherichia.  
 OC NCBI\_TaxID=562;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=X12 / MGI655;  
 RX MEDLINE=95334362; PubMed=7610040;  
 RA Burland V.D., Plunkett G. III, Sofia H.J., Daniels D.L.,  
 RA Blattner F.R.;  
 RA "Analysis of the Escherichia coli genome VI: DNA sequence of the  
 RT region from 92.8 through 100 minutes.";  
 RL Nucleic Acids Res. 23:2105-2119(1995).  
 CC [-] SIMILARITY: Contains 7 Keich repeats.  
 CC [-] SIMILARITY: STRONG, TO H.INFLUENZAE HI0148.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial



CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; U14003; AAA97206.1; ALT\_INIT.  
CC EMBL; AE000501; AAC77265.1; ALT\_INIT.  
CC EcoGene; EG12562; Yjht.  
CC InterPro; IPR006652; Kelch\_rep.  
CC Pfam; PF01344; Kelch; 2.  
KW Hypothetical protein; Kelch repeat; Repeat; Signal; Complete proteome.  
FT SIGNAL 1 19  
FT CHAIN 20 368  
FT REPEAT 40 84  
FT REPEAT 86 137  
FT REPEAT 139 173  
FT REPEAT 174 219  
FT REPEAT 222 265  
FT REPEAT 287 336  
FT REPEAT 338 367  
SQ SEQUENCE 368 AA; 39572 MW; 1194F392C51EA204 CRC64;  
  
Query Match 85.7%; Score 36; DB 1; Length 368;  
Best Local Similarity 87.5%; Pred. NO. 19;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
  
Oy 1 IGGEKAGG 8  
|||  
Db 341 IGGETAGG 348  
  
RESULT 5  
ID HS21\_MAIZE STANDARD; PRT; 161 AA.  
AC P24631;  
DT 01-MAR-1992 (Rel. 21, Created)  
DT 01-MAR-1992 (Rel. 21, Last sequence update)  
DT 01-JUN-1994 (Rel. 29, Last annotation update)  
DE 17.5 kDa class II heat shock protein.  
OS Zea mays (Maize)  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
OC PACAD clade; Panicoideae; Andropogoneae; Zea.  
OX NCBI\_TaxID=4577;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=cv. Ohio 43; TISSUE=Radicle;  
RA MEDLINE=91329703; PubMed=1714322;  
RA Goping I.S., Frappier J.R.H., Walden D.B., Atkinson B.G.;  
RT "Sequence, identification and characterization of cDNAs encoding two  
different members of the 18 kDa heat shock family of Zea mays L.";  
RL Plant Mol. Biol. 16:699-711(1991).  
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.  
CC -!- SIMILARITY: Belongs to the small heat shock protein (HSP20)  
family.  
CC -!- SIMILARITY: PLANTS HAVE FOUR CLASSES OF HSP20: CLASS I AND II  
WHICH ARE CYTOPLASMIC; CLASS III WHICH IS IN THE CHLOROPLAST  
AND CLASS IV WHICH IS IN THE ENDOMEMBRANE. THIS PROTEIN BELONGS  
TO CLASS II.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

KW Heat shock; Multigene family.  
SQ SEQUENCE 161 AA; 17511 MW; 3B60A8DD4396577C CRC64;  
  
Query Match 83.3%; Score 35; DB 1; Length 161;  
Best Local Similarity 85.7%; Pred. NO. 13;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
  
Oy 2 GGEKAGG 8  
|||  
Db 30 GGDKAGG 36  
  
RESULT 6  
FWDC\_METTH STANDARD; PRT; 270 AA.  
ID FWDC\_METTH  
AC O27600;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Tungsten-containing formylmethanofuran dehydrogenase II subunit C  
(EC 1.2.99.5).  
GN FWDC OR MTH1558.  
OS Methanobacterium thermoautotrophicum.  
OC Archaea; Euryarchaeota; Methanobacteria; Methanobacteriales;  
OC Methanobacteriaceae; Methanothermobacter.  
OX NCBI\_TaxID=187420;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Delta H;  
RX MEDLINE=98037514; PubMed=9371463;  
RA Smith D.R., Doucette-Stamm L.A., Deloughery C., Lee H.-M., Dubois J.,  
RA Aldredge T., Bashirzaden R., Blakely D., Cook R., Gilbert K.,  
RA Harrison D., Hoang L., Keagle P., Lum W., Pothier B., Qiu D.,  
RA Spadafora N., Vicare R., Wang Y., Wierzbowski J., Gibson R.,  
RA Jiواني N., Caruso A., Bush D., Safer H., Patwell D., Prabhakar S.,  
RA McDougall S., Shimer G., Goyal A., Pietrovski S., Church G.M.,  
RA Daniels C.J., Mao J.-I., Rice P., Noelling J., Reeve J.N.;  
RT "Complete genome sequence of Methanobacterium thermoautotrophicum  
deltaH: functional analysis and comparative genomics.";  
RL J. Bacteriol. 179:7135-7155(1997).  
CC -!- FUNCTION: CATALYZES THE REVERSIBLE OXIDATION OF CO(2) AND  
METHANOFURAN (MFR) TO N-FORMYLMETHANOFURAN (CHO-MFR). CAN ONLY  
OXIDISE FORMYLMETHANOFURAN. THIS ENZYME IS OXYGEN-LABILE.  
CC -!- CATALYTIC ACTIVITY: Formylmethanofuran + H(2)O + acceptor = CO(2)  
+ methanofuran + reduced acceptor.  
CC -!- COFACTOR: Tungsten.  
CC -!- ENZYME REGULATION: Not inactivated by cyanide.  
CC -!- PATHWAY: Methanogenesis; first step  
CC -!- SUBUNIT: THIS ENZYME IS COMPOSED OF SEVEN SUBUNITS FWDA (65 kDa),  
FWDB (53 kDa), FWDC (31 kDa), FWDD (15 kDa), FWDE, FWDF, AND FWDG.  
CC -!- INDUCTION: By growth on tungsten or molybdenum under anaerobic  
conditions.  
CC -!- SIMILARITY: Belongs to the fwdC/fmdC family.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; AE000916; AAB86032.1; -.

PIR; H69074; H69074.

InterPro; IPR002489; DUF14.

Pfam; PF01493; GKGXG; 1.

Oxidoreductase; Tungsten; Methanogenesis; Repeat; Complete proteome.

DOMAIN 80 213 7 X 13 AA REPEATS OF [GW]-X-X-M-X-X-G-

X-[IL]-X-[IV]-X-G.

REPEAT 80 92 1.

REPEAT 99 111 2.

REPEAT 118 130 3.

REPEAT 144 156 4.

```
FT REPEAT 163 175 5.
FT REPEAT 182 194 6.
FT REPEAT 201 213 7.
SQ SEQUENCE 270 AA; 28641 MW; 043A9FFBBA554D36 CRC64;

Query Match 83.3%; Score 35; DB 1; Length 270;
Best Local Similarity 75.0%; Pred. No. 22;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 IGGEKAGG 8
Db 200 VGGEMAGG 207

RESULT 7
FWDC METTM STANDARD; PRT; 270 AA.
AC Q59579; O08493;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Tungsten-containing formylmethanofuran dehydrogenase II subunit C
DE (EC 1.2.99.5).
GN FWDC.
OS Methanobacterium thermoautotrophicum (strain Marburg / DSM 2133).
OC Archaea; Euryarchaeota; Methanobacteria; Methanobacteriales;
OC Methanobacteriaceae; Methanothermobacter.
OX NCBI_TaxID=79929;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96163477; PubMed=8575452;
RA Hochheimer A., Schmitz R.A., Thauer R.K., Hedderich R.;
RT "The tungsten formylmethanofuran dehydrogenase from Methanobacterium
RT thermoautotrophicum contains sequence motifs characteristic for
RT enzymes containing molybdopterin dinucleotide.";
RL Eur. J. Biochem. 234:910-920(1995).
CC -1- FUNCTION: CATALYZES THE REVERSIBLE OXIDATION OF CO(2) AND
CC METHANOFURAN (MFR) TO N-FORMYLMETHANOFURAN (CHO-MFR). CAN ONLY
CC OXIDISE FORMYLMETHANOFURAN. THIS ENZYME IS OXYGEN-LABILE.
CC -1- CATALYTIC ACTIVITY: Formylmethanofuran + H(2)O + acceptor = CO(2)
CC + methanofuran + reduced acceptor.
CC -1- COFACTOR: Tungsten.
CC -1- PATHWAY: Methanogenesis; first step.
CC -1- SUBUNIT: THIS ENZYME IS COMPOSED OF SEVEN SUBUNITS FWDA (65 kDa),
CC FWDB (53 kDa), FWDC (31 kDa), FWDD (15 kDa), FWDE, FWDF, AND FWDG
CC (BY SIMILARITY).
CC -1- INDUCTION: By growth on tungsten or molybdenum under anaerobic
CC conditions.
CC -1- SIMILARITY: Belongs to the fwdc/fmdC family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X87970; CAA61214.1; -.
CC InterPro; IPR002489; DUF14.
CC Pfam; PF01493; GXGKG; 1.
CC Oxidoreductase; Tungsten; Methanogenesis; Repeat.
CC DOMAIN 80 213 7 X 13 AA REPEATS OF [GW]-X-X-M-X-X-G-
FT REPEAT 80 92 1.
FT REPEAT 99 111 2.
FT REPEAT 118 130 3.
FT REPEAT 144 156 4.
FT REPEAT 163 175 5.
FT REPEAT 182 194 6.
FT REPEAT 201 213 7.
SQ SEQUENCE 270 AA; 28554 MW; E0A369D2ACFEC46F CRC64;

Query Match 83.3%; Score 35; DB 1; Length 270;
Best Local Similarity 75.0%; Pred. No. 22;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 IGGEKAGG 8
Db 200 VGGEMAGG 207

RESULT 8
MERA ALCSP STANDARD; PRT; 559 AA.
AC P94188;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Mercuric reductase (EC 1.16.1.1) (Hg(II) reductase).
GN MERA.
OS Alcaligenes sp.
OG Plasmid IncHI2 pMER610.
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Alcaligenaceae; Alcaligenes.
OX NCBI_TaxID=512;
RN [1]
RP SEQUENCE FROM N.A.
RA Nikiforov V., Yurieva O., Kholodii G., Minakhin L., Gorlenko Z.,
RA Kalyaeva E., Mindlin S.;
RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: Resistance to Hg(2+) in bacteria appears to be governed
CC by a specialized system which includes mercuric reductase. MERA
CC protein is responsible for volatilizing mercury as Hg(0).
CC -1- CATALYTIC ACTIVITY: Hg + NADP(+) + H(+) = Hg(2+) + NADPH.
CC -1- COFACTOR: Binds 1 FAD per subunit (By similarity).
CC -1- SUBUNIT: Homodimer (By similarity).
CC -1- MISCELLANEOUS: The active site is a redox-active disulfide bond.
CC -1- SIMILARITY: Belongs to class-I pyridine nucleotide-disulfide
CC oxidoreductase family.
CC -1- SIMILARITY: Contains 1 HMA domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; Y08993; CAA70190.1; -.
CC HSP; Q04656; IAWO.
CC InterPro; IPR001327; FAD_pyr_redox.
CC InterPro; IPR006121; HeavyMe_transp.
CC InterPro; IPR000815; Hg_reductase.
CC InterPro; IPR006191; Metal_bind.
CC InterPro; IPR001100; Pyr_redox.
CC InterPro; IPR004099; pyr_redox_dim.
CC Pfam; PF00403; HMA_1;
CC Pfam; PF00070; Pyr_redox; 1.
CC Pfam; PF02852; Pyr_redox_dim; 1.
CC PRINTS; PR00368; FADPNR.
CC PRINTS; PR00945; HGRDTASE.
CC PRINTS; PR00411; PNRDRTASE1.
CC ProDom; PD000139; FAD_pyr_redox; 1.
CC PROSITE; PS01047; HMA_1; 1.
CC PROSITE; PS00846; HMA_2; 1.
CC PROSITE; PS00076; PYRIDINE_REDOX_1; 1.
CC Mercuric resistance; Oxidoreductase; Flavoprotein; FAD; NADP;
CC Mercury; Redox-active center; Metal-binding; Plasmid.
CC DOMAIN 1 65 HMA.
FT NP_BIND 126 134 FAD (ADP PART) (BY SIMILARITY).
FT DISULFID 134 139 REDOX-ACTIVE.
FT METAL 556 556 MERCURY (POTENTIAL).
FT METAL 557 557 MERCURY (POTENTIAL).
SQ SEQUENCE 559 AA; 58019 MW; 0603DDDB6B13CB519 CRC64;
```

```
Query Match 83.3%; Score 35; DB 1; Length 270;
Best Local Similarity 75.0%; Pred. No. 22;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 IGGEKAGG 8
Db 200 VGGEMAGG 207

RESULT 8
MERA ALCSP STANDARD; PRT; 559 AA.
AC P94188;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Mercuric reductase (EC 1.16.1.1) (Hg(II) reductase).
GN MERA.
OS Alcaligenes sp.
OG Plasmid IncHI2 pMER610.
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Alcaligenaceae; Alcaligenes.
OX NCBI_TaxID=512;
RN [1]
RP SEQUENCE FROM N.A.
RA Nikiforov V., Yurieva O., Kholodii G., Minakhin L., Gorlenko Z.,
RA Kalyaeva E., Mindlin S.;
RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: Resistance to Hg(2+) in bacteria appears to be governed
CC by a specialized system which includes mercuric reductase. MERA
CC protein is responsible for volatilizing mercury as Hg(0).
CC -1- CATALYTIC ACTIVITY: Hg + NADP(+) + H(+) = Hg(2+) + NADPH.
CC -1- COFACTOR: Binds 1 FAD per subunit (By similarity).
CC -1- SUBUNIT: Homodimer (By similarity).
CC -1- MISCELLANEOUS: The active site is a redox-active disulfide bond.
CC -1- SIMILARITY: Belongs to class-I pyridine nucleotide-disulfide
CC oxidoreductase family.
CC -1- SIMILARITY: Contains 1 HMA domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; Y08993; CAA70190.1; -.
CC HSP; Q04656; IAWO.
CC InterPro; IPR001327; FAD_pyr_redox.
CC InterPro; IPR006121; HeavyMe_transp.
CC InterPro; IPR000815; Hg_reductase.
CC InterPro; IPR006191; Metal_bind.
CC InterPro; IPR001100; Pyr_redox.
CC InterPro; IPR004099; pyr_redox_dim.
CC Pfam; PF00403; HMA_1;
CC Pfam; PF00070; Pyr_redox; 1.
CC Pfam; PF02852; Pyr_redox_dim; 1.
CC PRINTS; PR00368; FADPNR.
CC PRINTS; PR00945; HGRDTASE.
CC PRINTS; PR00411; PNRDRTASE1.
CC ProDom; PD000139; FAD_pyr_redox; 1.
CC PROSITE; PS01047; HMA_1; 1.
CC PROSITE; PS00846; HMA_2; 1.
CC PROSITE; PS00076; PYRIDINE_REDOX_1; 1.
CC Mercuric resistance; Oxidoreductase; Flavoprotein; FAD; NADP;
CC Mercury; Redox-active center; Metal-binding; Plasmid.
CC DOMAIN 1 65 HMA.
FT NP_BIND 126 134 FAD (ADP PART) (BY SIMILARITY).
FT DISULFID 134 139 REDOX-ACTIVE.
FT METAL 556 556 MERCURY (POTENTIAL).
FT METAL 557 557 MERCURY (POTENTIAL).
SQ SEQUENCE 559 AA; 58019 MW; 0603DDDB6B13CB519 CRC64;
```

Query Match 83.3%; Score 35; DB 1; Length 559;  
Best Local Similarity 85.7%; Pred. No. 44;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 GGEKAGG 8  
|||:||||  
DB 87 GGDKAGG 93

RESULT 9  
ID\_RL12 SCHPO STANDARD; PRT; 165 AA.  
AC 075000.  
DT 15-DEC-1998 (Rel. 37, Created)  
DT 15-DEC-1998 (Rel. 37, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE 60S ribosomal protein L12.  
GN (RPL12A OR SPOC31H12.04C) AND (RPL12B OR SPOC16C4.13C).

OS Schizosaccharomyces pombe (Fission yeast).  
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
OC Schizosaccharomycetes.  
OX NCBI\_TaxID=4896;  
RN [1]

RP SEQUENCE FROM N.A. (RPL12A AND RPL12B).  
RC STRAIN=972;

RX MEDLINE=21848401; PubMed=11859360;  
RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Stewart A.,

RA Sgourou J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,  
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,  
RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,  
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Fraser A.,  
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,  
RA James K., Jones L., Jones M., Leather S., McDonald S., Jagels K.,  
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,  
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,  
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,  
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,  
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,  
RA Woodward J., Volkart G., Aert R., Robben J., Grymonprez B.,  
RA Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,  
RA Gabel C., Fuchs M., Fritzc C., Holzer E., Moestl D., Hilbert H.,  
RA Borzym K., Langer I., Beck A., Leirach H., Reinhardt R., Pohl T.M.,  
RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,  
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,  
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,  
RA Lucas M., Rochet M., Gallard C., Talhada V.A., Garzon A., Thode G.,  
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,  
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,  
RA Cerutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,  
RA Spakovski G.V., Ussery D., Barrell B.G., Nurse P.;  
RT "The genome sequence of Schizosaccharomyces pombe."  
RL Nature 415:871-880(2002).

CC -!- FUNCTION: This protein binds directly to 26S ribosomal RNA (By  
similarity).

CC -!- MISCELLANEOUS: There are two genes for L12 in S.pombe.

CC -!- SIMILARITY: Belongs to the L12 family of ribosomal proteins.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; AL031824; CAA21221.1; -  
CC EMBL; AL031535; CAA20752.1; -  
CC F1R; T41103; T41103.  
CC HSSP; P29395; 1MMS.  
CC GeneDB SPombe; SPOC31H12.04c; -  
CC GeneDB SPombe; SPOC31H12.04c; -

DR InterPro; IPR000911; Ribosomal L11.  
DR Pfam; PF00298; Ribosomal L11; 1.  
DR Pfam; PF03946; Ribosomal L11\_N; 1.  
DR SMART; SM00649; RL11; 1.  
DR PROSITE; PS00359; RIBOSOMAL L11; 1.  
KW Ribosomal protein; RNA-binding; Multigene family.  
SQ SEQUENCE 165 AA; 17666 MW; FC77B46CD08F005 CRC64;

Query Match 81.0%; Score 34; DB 1; Length 165;  
Best Local Similarity 75.0%; Pred. No. 21;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 GGEKAGG 8  
:|||||  
DB 18 VGEVAGG 25

RESULT 10  
YGL2\_STRVR

ID YGL2\_STRVR STANDARD; PRT; 240 AA.

AC P19435;

DT 01-FEB-1991 (Rel. 17, Created)

DT 01-FEB-1991 (Rel. 17, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Hypothetical 23.9 kDa protein in glnII region (ORF2).

OS Streptomyces viridochromogenes.

OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;

OC Streptomycinae; Streptomycetaceae; Streptomyces.

OX NCBI\_TaxID=1938;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=ES2;

RX MEDLINE=90368592; PubMed=1975583;

RA Behrmann I., Hillemann D., Puchler A., Strauch E., Wohlleben W.;

RT "Overexpression of a Streptomyces viridochromogenes gene (glnII)

RT encoding a glutamine synthetase similar to those of eucaryotes confers

RT resistance against the antibiotic phosphinothricyl-alanyl-alanine.";

RL J. Bacteriol. 172:5326-5334(1990).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; X52842; CAA37027.1; -

DR PIR; B36724; B36724.

KW Hypothetical protein.

SQ SEQUENCE 240 AA; 23914 MW; 09DB6BEE681F7E2F CRC64;

Query Match 81.0%; Score 34; DB 1; Length 240;  
Best Local Similarity 85.7%; Pred. No. 30;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 GGEKAGG 8  
|||:||||  
DB 118 GGEQAGG 124

RESULT 11  
RIR1\_HSV11

ID\_RIR1\_HSV11 STANDARD; PRT; 1137 AA.

AC P08543;

DT 01-AUG-1988 (Rel. 08, Created)

DT 01-MAR-1989 (Rel. 10, Last sequence update)

DT 01-NOV-1995 (Rel. 32, Last annotation update)

DE Ribonucleoside-diphosphate reductase large chain (EC 1.17.4.1)

DE (Ribonucleotide reductase) (136 kDa subunit).

GN UR39.

OS Herpes simplex virus (type 1 / strain 17).

OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;

```
OC Alphaherpesvirinae; Simplexvirus.
OX NCBI_TaxID=10299;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=8274327; PubMed=2839594;
RA McGeoch D.J., Dallymple M.A., Davison A.J., Dolan A., Frame M.C.,
RA McNab D., Perry L.J., Scott J.E., Taylor P.;
RT "The complete DNA sequence of the long unique region in the genome of
RT herpes simplex virus type 1.";
RJ J. Gen. Virol. 69:1531-1574(1988).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=8217882; PubMed=2835765;
RA Nikas I., McLauchlan J., Davison A.J., Taylor W.R., Clements J.B.;
RT "Structural features of ribonucleotide reductase.";
RL Proteins 1:376-384(1986).
CC -!- FUNCTION: Provides the precursors necessary for DNA synthesis.
CC -!- CATALYTIC ACTIVITY: 2'-deoxyribonucleoside diphosphate + oxidized
CC thiorodoxin + H(2)O = ribonucleoside diphosphate + reduced
CC thiorodoxin.
CC -!- PATHWAY: DNA replication pathway; first step.
CC -!- SUBUNIT: Heterodimer of a large and a small chain.
CC -!- SIMILARITY: Belongs to the ribonucleoside diphosphate reductase
CC large chain family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X14112; CAA32314.1; -.
CC EMBL; M18410; AAA45805.1; -.
CC InterPro; IPR000788; Ribonucleo red.
CC Pfam; PF00317; ribonuc red lg; 1.
CC Pfam; PF02867; ribonuc_red_lgc; 1.
CC PRINTS; PR01183; RIBORDTASEM1.
CC PROSITE; PS00089; RIBORED LARG; 1.
CC Oxidoreductase; DNA replication; Early protein.
CC CONFLICT 70 N -> S (IN REF. 2).
CC CONFLICT 1034 1034 A -> P (IN REF. 2).
CC SEQUENCE 1137 AA; 124050 MW; 8A3777F4C22D8F85 CRC64;
Query Match 81.0%; Score 34; DB 1; Length 1137;
Best Local Similarity 62.5%; Pred. No. 1.3e+02;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
QY 1 IGGEKAGG 8
DB 17 VGGQEAGG 24
:|::|::|
RESULT 12
YD63 THEAC
ID -Y63_THEAC STANDARD; PRT; 312 AA.
AC QHHIG;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Putative HTH-type transcriptional regulatory protein Tal1363.
GN Tal1363.
OS Thermoplasma acidophilum.
OC Archaea; Euryarchaeota; Thermoplasmata; Thermoplasmatales;
OC Thermoplasmataceae; Thermoplasma.
OX NCBI_TaxID=2303;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=DSM 1728;
RA Ruepp A., Graml W., Santos-Martinez M.-L., Koretke K.K., Volker C.,
RA Mewes H.-W., Frishman D., Stocker S., Lupas A.N., Baumeister W.;
```

```
RT "The genome sequence of the thermoacidophilic scavenger Thermoplasma
RT acidophilum.";
RL Nature 407:508-513(2000).
CC -!- SIMILARITY: Contains 1 HTH cro/Cl-type DNA-binding domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AL445067; CAC12484.1; -.
CC HAMAP; MF_00584; 1.
CC InterPro; IPR001387; HTH_3.
CC Pfam; PF01381; HTH_3; 1.
CC SMART; SM00530; HTH_XRE; 1.
CC PROSITE; PS05943; HTH_CROCl; 1.
CC Hypothetical protein; Transcription regulation; DNA-binding;
CC Complete proteome.
CC DOMAIN 133 186 HTH CRO/Cl-TYPE.
CC DNA BIND 144 163 H-T-H MOTIF (POTENTIAL).
CC SEQUENCE 312 AA; 35272 MW; A0EC3FF0B3D72662 CRC64;
Query Match 78.6%; Score 33; DB 1; Length 312;
Best Local Similarity 87.5%; Pred. No. 59;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 IGGEKAGG 8
DB 77 IVGEKAGG 84
|::|::|
RESULT 13
TRMU VIBPA
ID -TRMU VIBPA STANDARD; PRT; 374 AA.
AC Q87OL9;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Probable tRNA (5-methylaminomethyl-2-thiouridylylate)-methyltransferase
DE (BC 2.1.1.61).
GN TRMU OR VP1130.
OS Vibrio parahaemolyticus.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC Vibrionaceae; Vibrio.
OX NCBI_TaxID=670;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=RIMD 2210633 / Serotype O3:K6;
RX MEDLINE=22508454; PubMed=12620739;
RA Makino K., Oshima K., Kurokawa K., Yokoyama K., Uda T., Tagomori K.,
RA Iijima Y., Najima M., Nakano M., Yamashita A., Kubota Y., Kimura S.,
RA Yasunaga T., Honda T., Shinagawa H., Hattori M., Iida T.;
RT "Genome sequence of Vibrio parahaemolyticus: a pathogenic mechanism
RT distinct from that of V. cholerae.";
RL Lancet 361:743-749(2003).
CC -!- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + tRNA = S-adenosyl-L-
CC homocysteine + tRNA containing 5-methylaminomethyl-2-
CC thiouridylylate.
CC -!- SIMILARITY: Belongs to the trmu family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AP005076; BAC59393.1; ALT_INIT.
```

```

DR HAMAP; MF 00144; -; 1.
DR InterPro; IPR004506; TrnU.
DR Pfam; PF03054; tRNA_Me_trans; 1.
DR TIGRFAMs; TIGR00420; trnU; 1.
KW Transferase; Methyltransferase; Complete proteome.
SQ SEQUENCE 374 AA; 42198 MW; 5BA02165BD84D5B5 CRC64;

Query Match 78.6%; Score 33; DB 1; Length 374;
Best Local Similarity 75.0%; Pred. No. 70;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 IGGEKAGG 8
Db 251 IGGKGGG 258

RESULT 15
TRMU VIBVU STANDARD; PRT; 374 AA.
AC Q8CWJ6;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Probable tRNA (5-methylaminomethyl-2-thiouridylyl)-methyltransferase
DE (EC 2.1.1.61).
GN TRMU OR VV12926.
OS Vibrio vulnificus.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC Vibrionaceae; Vibrio.
OX NCBI_TaxID=672;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CMCP6;
RA Rhee J.H., Kim S.Y., Chung S.S., Kim J.J., Moon Y.H., Jeong H.,
RA Choy H.E.;
RT "Complete genome sequence of Vibrio vulnificus CMCP6.";
RL Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.
CC -!- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + tRNA = S-adenosyl-L-
CC homocysteine + tRNA containing 5-methylaminomethyl-2-
CC thioridylate.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: Belongs to the trmU family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AE016806; AA011258.1; -.
CC HAMAP; MF 00144; -; 1.
CC InterPro; IPR004506; TrmU.
CC Pfam; PF03054; tRNA_Me_trans; 1.
CC TIGRFAMs; TIGR00420; trmU; 1.
KW Transferase; Methyltransferase; Complete proteome.
SQ SEQUENCE 374 AA; 42138 MW; F35A8671826EEC CRC64;

Query Match 78.6%; Score 33; DB 1; Length 374;
Best Local Similarity 75.0%; Pred. No. 70;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 IGGEKAGG 8
Db 251 IGGKGGG 258

RESULT 15
DP2L METAC STANDARD; PRT; 1145 AA.
AC Q8THG5;
DT 28-FEB-2003 (Rel. 41, Created)

```

```

DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE DNA polymerase II large subunit (EC 2.7.7.7) (Pol II).
GN POLC OR MA4552.
OS Methanosarcina acetivorans.
OC Archaea; Euryarchaeota; Methanomicrobia; Methanosarcinales;
OC Methanosarcinaceae; Methanosarcina.
OX NCBI_TaxID=22114;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C2A / ATCC 35395 / DSM 2834;
EX MEDLINE=21929760; PubMed=11932238;
RA Galagan J.E., Nusbaum C., Roy A., Endrizzi M.G., MacDonald P.,
RA FitzHugh W., Calvo S., Engels R., Smirnov S., Atnoor D., Brown A.,
RA Allen N., Naylor J., Stange-Thomann N., DeArellano K., Johnson R.,
RA Linton L., McEwan P., McKernan K., Talamas J., Tirrell A., Ye W.,
RA Zimmer A., Barber R.D., Cann I., Graham D.E., Grahame D.A., Guss A.M.,
RA Hedderich R., Ingram-Smith C., Kuettnher H.C., Krzycki J.A.,
RA Leigh J.A., Li W., Liu J., Mukhopadhyay B., Reeve J.N., Smith K.,
RA Springer T.A., Umayam L.A., White O., White R.H., de Macario E.C.,
RA Ferry J.G., Jarrell K.F., Jing H., Macario A.J.L., Paulsen I.,
RA Pritchett M., Sowers K.R., Swanson R.V., Zinder S.H., Lander E.,
RA Metcalf W.W., Birren B.;
RT "The genome of Methanosarcina acetivorans reveals extensive metabolic
RT and physiological diversity.";
RL Genome Res. 12:532-542(2002).
CC -!- FUNCTION: Possesses two activities: a DNA synthesis (polymerase)
CC and an exonucleolytic activity that degrades single stranded DNA
CC in the 3' to 5' direction. Has a template-primer preference which
CC is characteristic of a replicative DNA polymerase (By similarity).
CC -!- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate = N diphosphate
CC + {DNA}(N).
CC -!- CATALYTIC ACTIVITY: Degradation of single-stranded DNA. It acts
CC progressively in a 3'- to 5'-direction, releasing nucleoside 5'-
CC phosphates.
CC -!- SUBUNIT: Heterodimer of a large subunit and a small subunit (By
CC similarity).
CC -!- SIMILARITY: Belongs to the archaeal DNA polymerase II family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AE011176; AA007891.1; -.
CC HAMAP; MF 00324; -; 1.
CC InterPro; IPR004475; PolC_DP2.
CC Pfam; PF03833; PolC_DP2; 1.
CC TIGRFAMs; TIGR00354; polC; 1.
KW Transferase; DNA-directed DNA polymerase; DNA replication; Hydrolase;
KW Nuclease; Exonuclease; DNA-binding; Multifunctional enzyme;
KW Complete proteome.
SQ SEQUENCE 1145 AA; 127679 MW; 36C329A400C09E43 CRC64;

Query Match 78.6%; Score 33; DB 1; Length 1145;
Best Local Similarity 75.0%; Pred. No. 2.1e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 IGGEKAGG 8
Db 280 IGGAKGG 287

```

Search completed: September 18, 2004, 04:26:21  
Job time : 7.9e078 secs



GenCore version 5.1.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 18, 2004, 04:00:08 ; Search time 28.3922 Seconds  
(without alignments)  
88.903 Million cell updates/sec

Title: US-10-615-959-40

Perfect score: 42

Sequence: 1 IGGEXAGG 8

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL 25:\*

1: sp archaea:\*

2: sp bacteria:\*

3: sp fungi:\*

4: sp human:\*

5: sp invertebrate:\*

6: sp mammal:\*

7: sp mhc:\*

8: sp organelle:\*

9: sp phage:\*

10: sp plant:\*

11: sp rodent:\*

12: sp virus:\*

13: sp vertebrate:\*

14: sp unclassified:\*

15: sp virus:\*

16: sp bacteriophage:\*

17: sp archaea:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	42	100.0	1163	Q8N6U4	Q8N6U4 homo sapien
2	38	90.5	210	Q8PXL0	Q8PXL0 methanocarc
3	38	90.5	338	Q8BJU6	Q8BJU6 mus musculus
4	38	90.5	415	Q8SIV9	Q8SIV9 oryza sativ
5	38	90.5	460	Q7WY41	Q7WY41 pseudomonas
6	38	90.5	696	Q7XFP9	Q7XFP9 oryza sativ
7	38	90.5	1222	Q8K173	Q8K173 mus musculus
8	38	90.5	1464	Q8BK72	Q8BK72 mus musculus
9	38	90.5	1464	Q8BLW4	Q8BLW4 mus musculus
10	38	90.5	1464	Q7TT32	Q7TT32 mus musculus
11	37	88.1	378	Q93IK5	Q93IK5 vibrio sp.
12	37	88.1	476	Q8ZL18	Q8ZL18 salmonella
13	36	85.7	167	O15729	O15729 entamoeba h
14	36	85.7	322	O06441	O06441 rhodococcus
15	36	85.7	338	Q82R45	Q82R45 streptomyces
16	36	85.7	397	Q83IN1	Q83IN1 shigella fl

#### ALIGNMENTS

##### RESULT 1

Q8N6U4  
ID Q8N6U4 PRELIMINARY; PRT; 1163 AA.  
AC Q8N6U4  
DT 01-OCT-2002 (Tremblrel. 22, Created)  
DT 01-OCT-2002 (Tremblrel. 22, Last sequence update)  
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)  
DE Similar to collagen, type III, alpha 1 (Ehlers-Danlos syndrome type IV, autosomal dominant).  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Liver;  
RA Strausberg R.;  
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.  
DR GO; GO:0005581; C:collagen; IEA.  
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.  
DR InterPro; IPR008160; Collagen.  
DR InterPro; IPR002181; Fibrinogen C.  
DR InterPro; IPR000885; Fib collagen\_C.  
DR Pfam; PF01410; COLFI; 1.  
DR Pfam; PF01391; Collagen; 13.  
DR ProDom; PDOM02078; Fib\_collagen\_C; 1.  
DR SMART; SM00038; COLFI; 1.  
DR SMART; SM00214; VWC; 1.  
DR PROSITE; PS01208; VWC\_1; 1.  
DR PROSITE; PS0184; VWC\_2; 1.  
KW Collagen.  
SQ SEQUENCE 1163 AA; 111899 MW; 9E0C6BE1E94D6357 CRC64;

Query Match

Best Local Similarity 100.0%; Score 42; DB 4; Length 1163;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

—



DR	GO: 0005201; F:extracellular matrix structural constituent; IEA.
DR	InterPro; IPR008161; Clg_helix.
DR	InterPro; IPR008160; Collagen.
DR	InterPro; IPR002181; Fibrinogen_C.
DR	InterPro; IPR000885; Fib_collagen_C.
DR	Pfam; PF01410; COLFI; 1.
DR	Pfam; PF01391; Collagen; 16.
DR	ProDom; PD000007; Clg_helix; 1.
DR	ProDom; PD002078; Fib_collagen_C; 1.
DR	SMART; SM00038; COLFI; 1.
KW	Hypothetical protein; Collagen.
FT	NON_TER 1 1
SQ	SEQUENCE 1222 AA; 115140 MW; A409CA00D8D2765E4 CRC64;
Query Match 90.5%; Score 38; DB 11; Length 1222;	
Best Local Similarity 75.0%; Pred. No. 3e+02; Indels 0; Gaps 0;	
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;	
Qy	1 IGGEKAGG 8
Db	:    :
	964 VGGEKSGG 971
RESULT 8	
Q8BKX2	
ID	Q8BKX2 PRELIMINARY; PRT; 1464 AA.
AC	Q8BKX2;
DT	01-MAR-2003 (TREMELrel. 23, Created)
DT	01-MAR-2003 (TREMELrel. 23, Last sequence update)
DT	01-OCT-2003 (TREMELrel. 25, Last annotation update)
DE	Collagen alpha 1.
GN	COL3A1.
OS	Mus musculus (Mouse).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX	NCBI_TaxID=10090;
RN	[1]
RC	SEQUENCE FROM N.A.
RP	STRAIN=C57BL/6J; TISSUE=Head;
RX	MEDLINE=22354683; PubMed=12466851;
RA	The FANTOM Consortium,
RA	the RIKEN Genome Exploration Research Group Phase I & II Team;
RT	"Analysis of the mouse transcriptome based on functional annotation of
RT	60,770 full-length cDNAs.";
RL	Nature 420:563-573(2002).
DR	EMBL; AK048546; BAC33370.1; -.
DR	MED; MGI:88453; Col3a1.
DR	GO: 0005581; C:collagen; IEA.
DR	GO: 0005201; F:extracellular matrix structural constituent; IEA.
DR	InterPro; IPR008161; Clg_helix.
DR	InterPro; IPR008160; Collagen.
DR	InterPro; IPR002181; Fibrinogen_C.
DR	InterPro; IPR000885; Fib_collagen_C.
DR	InterPro; IPR001007; VWFC.
DR	Pfam; PF01410; COLFI; 1.
DR	Pfam; PF01391; Collagen; 18.
DR	ProDom; PD000007; Clg_helix; 1.
DR	ProDom; PD002078; Fib_collagen_C; 1.
DR	SMART; SM00038; COLFI; 1.
DR	SMART; SM00214; VMC; 1.
DR	PROSITE; PS01208; VWFC 1; 1.
DR	PROSITE; PS0184; VWFC 2; 1.
DR	SEQUENCE 1464 AA; 138947 MW; 1E4ED9539EF42C12 CRC64;
Query Match 90.5%; Score 38; DB 11; Length 1464;	
Best Local Similarity 75.0%; Pred. No. 3.6e+02; Indels 0; Gaps 0;	
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;	
Qy	1 IGGEKAGG 8
Db	:    :
	1206 VGGEKSGG 1213

```

RESULT 9
Q8BLW4 Q8BLW4 PRELIMINARY; PRT; 1464 AA.
AC Q8BLW4;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Collagen alpha 1.
GN COL3A1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Aorta, and Vein;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium.
RA The RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs."
RL Nature 420:563-573(2002).
DR EMBL; AK041115; BAC30826.1; -.
DR MGD; MGI:88453; Col3a1.
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR InterPro; IPR008161; Clg_helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR002181; Fibrinogen_C.
DR InterPro; IPR000885; Fib_collagen_C.
DR InterPro; IPR001007; VWF_C.
DR Pfam; PF01410; COLF1; 1.
DR Pfam; PF01391; Collagen; 18.
DR ProDom; PD000007; Clg_helix; 1.
DR ProDom; PD002078; Fib_collagen_C; 1.
DR SMART; SM00038; COLF1; 1.
DR SMART; SM00214; VWC; 1.
DR PROSITE; PS01208; VWF_C_1; 1.
DR PROSITE; PS01018; VWF_C_2; 1.
SQ SEQUENCE 1464 AA; 138939 MW; 91F3246D90818449 CRC64;

Query Match 90.5%; Score 38; DB 11; Length 1464;
Best Local Similarity 75.0%; Pred. No. 3.6e+02;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 IGGEKAGG 8
Db 1206 VGGEKSGG 1213
:|||||:

RESULT 10
Q7TT32 Q7TT32 PRELIMINARY; PRT; 1464 AA.
AC Q7TT32;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Brain;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,

Query Match 90.5%; Score 38; DB 11; Length 1464;
Best Local Similarity 75.0%; Pred. No. 3.6e+02;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 IGGEKAGG 8
Db 1206 VGGEKSGG 1213
:|||||:

RESULT 11
Q93IK5 Q93IK5 PRELIMINARY; PRT; 378 AA.
AC Q93IK5;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein (Outer membrane protein).
OS Vibrio sp. CH-291.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC Vibrionaceae; Vibrio.
OX NCBI_TaxID=161725;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CH-291;
RA Jores J., Lewin A., Appel B.;
RT "Cloning of a hemolysin encoding region of a Vibrio species.";
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: OUTER MEMBRANE (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE OMP25 / ROPB FAMILY.
DR EMBL; AJ314791; CAC40971.1; -.
DR GO; GO:0009279; C:external outer membrane (sensu Gram-negativ. .; IEA.
DR GO; GO:0019867; C:outer membrane; IEA.
DR InterPro; IPR006664; Bac_OmpA.
DR InterPro; IPR006665; OmpA/MotB.
DR InterPro; IPR000498; OmpA_tmam.
DR Pfam; PF00691; OmpA; 1.
DR Pfam; PF01389; OmpA membrane; 1.
DR PRINTS; PR01021; OMPADOMAIN.
DR ProDom; PD000930; OmpA/MotB; 1.
DR Hypothetical protein; Antigen; Outer membrane.
SQ SEQUENCE 378 AA; 40886 MW; A5E50FAC4D8814C7 CRC64;

Query Match 88.1%; Score 37; DB 2; Length 378;
Best Local Similarity 75.0%; Pred. No. 1.3e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 IGGEKAGG 8
Db 196 VGGEKTCG 193
:|||||:

```

```

RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Boeak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hate S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smallus D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Brain;
RA Strausberg R.;
RL Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC052398; AAH52398.1; -.
KW Hypothetical protein.
SQ SEQUENCE 1464 AA; 138971 MW; 2B38BC27AF21590B CRC64;

Query Match 90.5%; Score 38; DB 11; Length 1464;
Best Local Similarity 75.0%; Pred. No. 3.6e+02;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 IGGEKAGG 8
Db 1206 VGGEKSGG 1213
:|||||:

```

## RESULT 12

Q82L18  
ID Q82L18 PRELIMINARY; PRT; 476 AA.  
AC Q82L18;  
DT 01-MAR-2002 (TrEMBLrel. 20, Created)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
DE Hypothetical protein STV4577.  
GN STV4577 OR T4274.  
OS Salmonella typhi.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Salmonella.  
OX NCBI\_TaxID=601;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CT18;  
RX MEDLINE=21534947; PubMed=11677608;  
RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,  
RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebaihia M.,  
RA Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,  
RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,  
RA Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,  
RA Krogn A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,  
RA Quail M., Rutherford K., Leather S., Skelton J., Stevens K.,  
RA Whitehead S., Barrell B.G.;  
RT "Complete genome sequence of a multiple drug resistant Salmonella  
enterica serovar Typhi CT18.";  
RL Nature 413:848-852(2001).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Ty2 / ATCC 700931;  
RX MEDLINE=22531367; PubMed=12644504;  
RA Deng W., Liou S.-R., Plunkett G. III, Mayhew G.F., Rose D.J.,  
RA Burland V., Kodoyanni V., Schwartz D.C., Blattner F.R.;  
RT "Comparative genomics of Salmonella enterica serovar Typhi strains Ty2  
and CT18.";  
RL J. Bacteriol. 185:2330-2337(2003).  
DR EMBL; AL627282; CAD09352.1; -;  
DR EMBL; AF016848; AAO71732.1; -;  
KW Hypothetical protein; Complete proteome.

SEQUENCE 476 AA; 49873 MW; 351CBF96D8B7A8D7 CRC64;

Query Match 88.1%; Score 37; DB 16; Length 476;  
Best Local Similarity 75.0%; Pred. No. 1.7e+02;  
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 IGGEKAGG 8

Db 211 VGGKAGG 218

## RESULT 13

O15729  
ID O15729 PRELIMINARY; PRT; 167 AA.  
AC O15729;  
DT 01-JAN-1998 (TrEMBLrel. 05, Created)  
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)  
DE Cyclophilin (EC 5.2.1.8) (Peptidyl-prolyl cis-trans isomerase)  
DE (Ppiase) (Rotamase).  
GN CYP1.  
OC Entamoeba histolytica.  
OC Eukaryota; Entamoebidae; Entamoeba.  
OX NCBI\_TaxID=5759;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=HMI:IMSS;  
RX MEDLINE=20243663; PubMed=10779598;  
RA Ostoa-Saloma P., Carrero J.C., Petrossian P., Herion P., Landa A.,  
RA Laclette J.P.;  
RT "Cloning, characterization and functional expression of a cyclophilin  
Mol. Biochem. Parasitol. 107:219-225(2000)."

## [2]

RN SEQUENCE FROM N.A.  
RP Caballero-Franco C., Petrosyan P., Carrero-Sanchez J.C.,  
RA De la Torre P., Laclette J.P., Ostoa-Saloma P.;  
RT "Entamoeba histolytica nucleotide sequence containing the gene for a  
small heat shock protein, the CYP1 cyclophilin gene and a truncated  
ORF for a putative serine/threonine protein kinase.";  
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: PPIASES ACCELERATE THE FOLDING OF PROTEINS (BY  
SIMILARITY).  
CC -!- CATALYTIC ACTIVITY: PEPTIDYLPROLINE (OMEGA=180) = PEPTIDYLPROLINE  
(OMEGA=0).  
CC -!- SIMILARITY: BELONGS TO THE CYCLOPHILIN-TYPE PPIASE FAMILY.  
DR EMBL; AF017993; AAB86601.1; -;  
DR EMBL; AF498649; AAM21054.1; -;  
DR HSSP; P05092; 2CPL.  
DR GO; GO:0042027; F:cyclophilin-type peptidyl-prolyl cis-trans . . .; IEA.  
DR GO; GO:0004600; F:cyclophilin; IEA.  
DR GO; GO:0030053; F:FK506-sensitive peptidyl-prolyl cis-trans i . . .; IEA.  
DR GO; GO:0006853; F:isomerase activity; IEA.  
DR GO; GO:0006457; F:protein folding; IEA.  
DR InterPro; IPR002130; CSA\_PPIase.  
DR Pfam; PF00160; pro isomerase; 1.  
DR PRINTS; PR00153; CSAPPISMRASE.  
DR PROSITE; PS00170; CSA\_PPIASE\_1; 1.  
DR PROSITE; PS00072; CSA\_PPIASE\_2; 1.  
KW Isomerase; Rotamase.  
SQ SEQUENCE 167 AA; 18139 MW; DBA23FE5BB3290B2 CRC64;

Query Match 85.7%; Score 36; DB 5; Length 167;  
Best Local Similarity 100.0%; Pred. No. 84;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IGGEKAGG 7

Db 12 IGGEKAGG 18

## RESULT 14

O06441  
ID O06441 PRELIMINARY; PRT; 322 AA.  
AC O06441;  
DT 01-JUL-1997 (TrEMBLrel. 04, Created)  
DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Heroin esterase.  
GN HER.  
OS Rhodococcus sp.  
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
OC Corynebacterineae; Nocardiaceae; Rhodococcus.  
OX NCBI\_TaxID=1831;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=H1;  
RX MEDLINE=97288094; PubMed=9143135;  
RA Rathbone D.A., Holt P.J., Lowe C.R., Bruce N.C.;  
RT "Molecular analysis of the Rhodococcus sp. strain H1 her gene and  
characterization of its product, a heroin esterase, expressed in  
Escherichia coli.";  
RL Appl. Environ. Microbiol. 63:2062-2066(1997).  
CC -!- SIMILARITY: BELONGS TO THE TYPE-B CARBOXYLESTERASE/LIPASE FAMILY.  
DR EMBL; U70619; AAC45283.1; -;  
DR GO; GO:0016787; F:Hydrolase activity; IEA.  
DR InterPro; IPR002018; CarbesteraseB.  
DR InterPro; IPR000379; Ser esterase.  
DR Pfam; PF00135; COesterase; 1.  
KW Hydrolase.  
SQ SEQUENCE 322 AA; 34239 MW; 93B01D5C760B4A1E CRC64;

Query Match 85.7%; Score 36; DB 2; Length 322;  
Best Local Similarity 75.0%; Pred. No. 1.7e+02;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Mon Sep 20 11:05:44 2004

QY 1 IGGEKAGG 8  
:|||||  
Db 156 VGESAGG 163

RESULT 15

Q82R45 PRELIMINARY; PRT; 338 AA.  
AC Q82R45;  
DT 01-JUN-2003 (TrEMBLrel. 24, Created)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Putative arylesterase/monooxygenase.  
GN SAV239.  
OS Streptomyces avermitilis.  
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
OC Streptomycineae; Streptomycetaceae; Streptomyces.  
OX NCBI\_TaxID=33903;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;  
RX MEDLINE=21477403; PubMed=11572948;  
RA Omura S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,  
RA Shinose M., Takahashi Y., Horikawa H., Nakazawa H., Osone T.,  
RA Kikuchi H., Shiba T., Sakaki Y., Hattori M.,  
RT "Genome sequence of an industrial microorganism Streptomyces  
RT avermitilis; deducing the ability of producing secondary  
RT metabolites."  
RL Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220(2001).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;  
RX MEDLINE=22608306; PubMed=12692562;  
RA Ikeda H., Ishikawa J., Hanamoto A., Shinose M., Kikuchi H., Shiba T.,  
RA Sakaki Y., Hattori M., Omura S.;  
RT "Complete genome sequence and comparative analysis of the industrial  
RT microorganism Streptomyces avermitilis."  
RL Nat. Biotechnol. 21:526-531(2003).  
DR EMBL; AP005022; BAC68008.1; -.  
DR GO; GO:0003624; F:catalytic activity; IEA.  
DR InterPro; IPR000379; Ser\_estra.  
KW Complete proteome.  
SQ SEQUENCE 338 AA; 36242 MW; EA33486345FD82EE CRC64;

Query Match 85.7%; Score 36; DB 16; Length 338;  
Best Local Similarity 75.0%; Pred.No. 1.8e+02;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 IGGEKAGG 8  
:|||||  
Db 177 VGESAGG 184

Search completed: September 18, 2004, 04:29:37  
Job time : 31.3922 secs

GenCore version 5.1.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 18, 2004, 03:47:39 ; Search time 48.1765 Seconds  
(without alignments)  
52.784 Million cell updates/sec

Title: US-10-615-959-41

Perfect score: 48

Sequence: 1 IGGEXAGGF 9

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A.Geneseq 29Jan04:\*

1: geneseqp1980s:\*

2: geneseqp1990s:\*

3: geneseqp2000s:\*

4: geneseqp2001s:\*

5: geneseqp2002s:\*

6: geneseqp2003as:\*

7: geneseqp2003bs:\*

8: geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	48	100.0	9	4	AAB61741 Human typ
2	48	100.0	12	4	AAB61743 Human typ
3	48	100.0	25	7	AAB38632 Human C-t
4	48	100.0	30	2	AAR92068 Recombina
5	48	100.0	262	7	AAB38640 Human col
6	48	100.0	270	2	AAY07375 Procollag
7	48	100.0	293	6	ABU70813 Human adi
8	48	100.0	309	4	AAB23675 Novel hum
9	48	100.0	309	4	ABG60248 Human ova
10	48	100.0	309	5	ABG61719 Novel ova
11	48	100.0	339	6	ABU70735 Human adi
12	48	100.0	623	2	AAB12843 Pro-alpha
13	48	100.0	626	2	AAB12842 Truncated
14	48	100.0	842	7	ABE08475 Novel pro
15	48	100.0	1466	4	ABE08475 Collagen
16	48	100.0	1466	5	ABE08475 Collagen
17	48	100.0	1466	5	ABE08475 Collagen
18	48	100.0	1466	6	ABU54454 Human tum
19	48	100.0	1466	6	ABU54454 Breast ca
20	48	100.0	1469	4	ABG15191 Novel hum
21	48	100.0	1470	7	ABE09399 Novel pro
22	47	97.9	1726	6	ABR42661 Decorin-m
23	47	97.5	1466	4	ABE02537 Porcine a
24	42	87.5	8	4	AAB61740 Human typ
25	42	87.5	11	3	AAY68425 Type III
			11	3	AAY68436 Type III

26	42	87.5	11	3	AAY82167
27	42	87.5	11	4	AAB61742
28	42	87.5	11	4	AAB51275 Collagen
29	42	87.5	1078	2	AAR71704 Collagen
30	42	87.5	1078	3	AAY96125 Collagen
31	42	87.5	1078	5	AAB16478 Human col
32	42	87.5	1078	5	ABB80736 Collagen
33	42	87.5	1078	5	ABB09628 Amino aci
34	41	85.4	691	6	ABU19644 Protein e
35	41	85.4	1466	4	AAB02534 Bovine al
36	41	85.4	1466	4	AAB02533 Bovine al
37	40	83.3	133	4	AAB00933 Human pol
38	38	79.2	51	5	ABP33968 Human ORF
39	38	79.2	142	3	AAG40971 Zea mays
40	38	79.2	674	2	AAY29125 Amino aci
41	37	77.1	481	6	ABM70142 Phototrab
42	35	72.9	10	2	AAB27523 Murine in
43	35	72.9	57	7	ADB81619 Human ova
44	35	72.9	83	4	AAB03625 Human pol
45	35	72.9	128	2	AAY20721 Human neu

## ALIGNMENTS

RESULT 1  
AAB61741  
ID AAB61741 standard; peptide; 9 AA.  
XX  
AC AAB61741;  
XX  
DT 20-APR-2001 (first entry)  
XX  
DE Human type III collagen carboxy-telopeptide fragment.  
XX  
KW Type II collagen; immunoassay; cartilage; telopeptide; human;  
KW type III collagen.  
XX  
OS Synthetic.  
OS Homo sapiens.  
PN WO200079284-A1.  
XX  
PD 28-DEC-2000.  
XX  
PF 10-DEC-1999; 99WO-US029357.  
XX  
PR 17-JUN-1999; 99US-00335098.  
PR 29-JUN-1999; 99US-0141574P.  
PR 02-JUL-1999; 99US-0142274P.  
PR 07-JUL-1999; 99US-0142675P.  
PR 30-AUG-1999; 99US-00385740.  
XX  
PA (WASH-) WASHINGTON RES FOUND.  
XX  
PI Eyre DR;  
XX  
WPI; 2001-146859/15.  
XX  
PT Assay for detecting cross-linked telopeptide analytes indicative of type II collagen resorption in vivo in a body fluid sample, comprises contacting the sample with an antibody which binds to the analyte.  
XX  
PS Disclosure; Page 15; 34pp; English.  
XX  
CC The invention relates to immunoassays for measuring type II collagen (cartilage) resorption in vivo. The method of analysing a body fluid sample for the presence of an analyte indicative of a physiological condition, involves contacting the body fluid sample with an antibody (Ab) which binds to the analyte, detecting binding of Ab in the body fluid sample, and correlating any detected binding to the physiological condition. The analysis is useful for measuring type II collagen (cartilage) resorption in vivo, for distinguishing between resorption of

CC non-mineralized and mineralized cartilage, and for measuring total  
 CC cartilage resorption in vivo. The present sequence represents a linear  
 CC synthetic peptide based on the carboxy-terminal telopeptide sequence of  
 CC human collagen type III  
 XX human collagen type III  
 SQ Sequence 9 AA;

Query Match 100.0%; Score 48; DB 4; Length 9;  
 Best Local Similarity 100.0%; Pred. No. 1.4e+06;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 IGGEKAGGF 9  
 |||||  
 Db 1 IGGEKAGGF 9

RESULT 2  
 AAB61743  
 ID AAB61743 standard; peptide; 12 AA.

XX AAB61743;

DT 20-APR-2001 (first entry)

XX Human type III collagen carboxy-telopeptide fragment.

XX Type II collagen; immunoassay; cartilage; telopeptide; human;  
 KW type III collagen.

XX Synthetic.

OS Homo sapiens.

XX WO200079284-A1.

XX 28-DEC-2000.

XX 10-DEC-1999; 99WO-US029357.

XX 17-JUN-1999; 99US-00335098.

PR 29-JUN-1999; 99US-0141574P.

PR 02-JUL-1999; 99US-0142274P.

PR 07-JUL-1999; 99US-0142675P.

XX 30-AUG-1999; 99US-00385740.

XX (WASH-) WASHINGTON RES FOUND.

XX Eyre DR;

XX WPI; 2001-146859/15.

XX Assay for detecting cross-linked telopeptide analytes indicative of type  
 PT II collagen resorption in vivo in a body fluid sample, comprises  
 PT contacting the sample with an antibody which binds to the analyte.  
 XX Disclosure; Page 15; 34pp; English.

XX The invention relates to immunoassays for measuring type II collagen  
 CC (cartilage) resorption in vivo. The method of analysing a body fluid  
 CC sample for the presence of an analyte indicative of a physiological  
 CC condition, involves contacting the body fluid sample with an antibody  
 CC (Ab) which binds to the analyte, detecting binding of Ab in the body  
 CC fluid sample, and correlating any detected binding to the physiological  
 CC condition. The analysis is useful for measuring type II collagen  
 CC (cartilage) resorption in vivo, for distinguishing between resorption of  
 CC non-mineralized and mineralized cartilage, and for measuring total  
 CC cartilage resorption in vivo. The present sequence represents a linear  
 CC synthetic peptide based on the carboxy-terminal telopeptide sequence of  
 CC human collagen type III

XX Sequence 12 AA;

Query Match 100.0%; Score 48; DB 4; Length 12;  
 Best Local Similarity 100.0%; Pred. No. 0.11;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 1 IGGEKAGGF 9  
 |||||  
 Db 4 IGGEKAGGF 12

RESULT 3  
 AAE38632  
 ID AAE38632 standard; peptide; 25 AA.

XX AAE38632;

DT 04-DEC-2003 (first entry)

XX Human C-terminal telopeptide alpha 1 #4.

XX Human; therapy; osteoporosis; left-ventricular hypertrophy; scleroderma;  
 KW liver cirrhosis; congestive heart failure; multiple sclerosis; fibrosis;  
 KW coal-miner's pneumoconiosis; muscle-wasting syndrome; diabetes mellitus;  
 KW rheumatoid arthritis; Alzheimer's disease; muscular dystrophy; cancer;  
 KW telopeptide alpha.

OS Homo sapiens.

PN WO2003068919-A2.

XX 21-AUG-2003.

XX 12-FEB-2003; 2003WO-US004183.

XX 12-FEB-2002; 2002US-0356008P.

XX (REGC ) UNIV CALIFORNIA.

XX Hellerstein MK;

XX WPI; 2003-689661/65.

XX Determining rate of biosynthesis or breakdown of inaccessible biological  
 PT molecules, useful e.g. for diagnosis or monitoring treatment, by  
 PT administering labeled precursor.  
 XX Claim 24; Page 99; 105pp; English.

XX The invention relates to a method of determining the rate of biosynthesis  
 CC or breakdown of at least one inaccessible biological molecule in a  
 CC subject. The method is useful for diagnosis or monitoring and treatment  
 CC of diseases associated with an altered rate of biosynthesis/breakdown of  
 CC an isotopically labelled precursor molecule, specifically osteoporosis;  
 CC left-ventricular hypertrophy; liver cirrhosis or fibrosis; congestive  
 CC heart failure; scleroderma; coal-miner's pneumoconiosis; cardiac or lung  
 CC fibrosis; Alzheimer's disease; multiple sclerosis; rheumatoid arthritis;  
 CC diabetes mellitus; muscle-wasting syndromes; muscular dystrophy; athletic  
 CC training and cancer. The method is also useful for screening candidate  
 CC gene or protein targets, phenotypic/human validation studies on potential  
 CC drugs, drug mechanism studies and determining the risk of developing the  
 CC disease. The present sequence is human C-terminal telopeptide alpha 1.  
 CC This sequence is used to illustrate the method of the invention

XX Sequence 25 AA;

Query Match 100.0%; Score 48; DB 7; Length 25;

Best Local Similarity 100.0%; Pred. No. 0.23;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 IGGEKAGGF 9  
 |||||  
 Db 12 IGGEKAGGF 20

RESULT 4  
 AAR92068



Query Match 100.0%; Score 48; DB 7; Length 262;  
 Best Local Similarity 100.0%; Pred. No. 2.6;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IGGEKAGGF 9  
 DB 4 IGGEKAGGF 12

RESULT 6  
 AAY07375  
 ID AAY07375 standard; protein; 270 AA.  
 XX  
 AC AAY07375;  
 XX  
 DT 16-JUL-1999 (first entry)  
 XX  
 DE Procollagen-III-C-terminal propeptide.  
 XX  
 KW Alcoholic cirrhosis; biliary cirrhosis; hepatitis; schistosomiasis;  
 KW cardiac fibrosis; Crohn's disease; diabetic nephropathy; collagen;  
 KW fibril; procollagen-III-C-terminal propeptide; immunoassay; diagnosis;  
 KW glucocorticosteroid; antibody.  
 XX  
 OS Homo sapiens.  
 XX  
 PN EP913692-A1.  
 XX  
 PD 06-MAY-1999.  
 XX  
 PF 31-OCT-1997; 97EP-00119018.  
 XX  
 PR 31-OCT-1997; 97EP-00119018.  
 XX  
 PA (FARB ) BAYER AG.  
 XX  
 PI Burchardt ER, Kroll W, Neumann R, Schroeder W;  
 XX  
 DR WPI; 1999-256748/22.  
 DR N-PSDB; AAX57501.  
 XX  
 PT New immunoassay for procollagen-III-C-terminal propeptide (PIIICP) useful  
 for diagnosing diseases such as alcoholic cirrhosis, hepatitis and  
 Duchenne's muscular dystrophy.  
 XX  
 PS Disclosure; Fig 1; 28pp; English.

XX  
 CC Diseases such as alcoholic cirrhosis, biliary cirrhosis, hepatitis,  
 CC schistosomiasis, cardiac fibrosis, Crohn's disease, diabetic nephropathy  
 CC and fibroses caused by surgery result from an inappropriate production of  
 CC collagen fibrils. Collagen III is synthesised as a preproprotein which is  
 CC modified by post-translation modification. The mature collagen III  
 CC fibrils are generated by cleavage of a propeptide from the C-terminus  
 CC followed by cleavage of a fragment from the N-terminus. This sequence  
 CC corresponds to the procollagen III C-terminal propeptide (PIIICP). As the  
 CC fibrils are formed, the PIIICP propeptide is released into the  
 CC surrounding tissue fluids. The invention relates to an immunoassay to  
 CC determine PIIICP concentrations and thus diagnose the above diseases and  
 CC others which result from inappropriate collagen deposition. Different  
 CC fibrotic diseases may show different levels of PIIICP in the serum so the  
 CC immunoassay may also be used to determine which disease a particular  
 CC patient is suffering from. The new immunoassay may also be used to  
 CC measure the rate of collagen synthesis in patients being treated with  
 CC glucocorticosteroids. In addition the new antibodies may be used in  
 CC immunohistochemical staining of cryostat and paraffin sections to assess  
 CC collagen synthesis in tissue samples from patients suspected of having  
 CC fibrotic disease  
 XX  
 SQ Sequence 270 AA;

Query Match 100.0%; Score 48; DB 2; Length 270;  
 Best Local Similarity 100.0%; Pred. No. 2.7;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 IGGEKAGGF 9  
 DB 12 IGGEKAGGF 20

RESULT 7  
 ABU70813  
 ID ABU70813 standard; protein; 293 AA.  
 XX  
 AC ABU70813;  
 XX  
 DT 10-JUN-2003 (first entry)  
 XX  
 DE Human adipocyte Selected Interacting domain, SID, #444.

XX  
 KW Human; prey; adipocyte; SID; selected interacting domain; anorectic;  
 KW anti-diabetic; protein-protein interaction; diabetes;  
 KW yeast 2-hybrid assay; metabolic disorder; obesity.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200286122-A2.  
 XX  
 PD 31-OCT-2002.  
 XX  
 PF 14-MAR-2002; 2002WO-BF003768.  
 XX  
 PR 14-MAR-2001; 2001US-0275734P.  
 XX  
 PA (HYBR-) HYBRIGENICS.  
 XX  
 PI Legrain P, Daviet L;  
 XX  
 DR WPI; 2003-103412/09.  
 DR N-PSDB; ACA57357.

XX  
 PT New complex between two interacting proteins in adipocyte cells, useful  
 for identifying selected interacting domains that modulate protein  
 interactions, or for preventing or treating metabolic disorders such as  
 obesity or diabetes.  
 XX  
 PS Claim 6; Page 258-259; 382pp; English.

XX  
 CC The invention relates to a complex between two interacting proteins in  
 CC adipocyte cells, given in the specification. The proteins are identified  
 CC by selecting a bait protein from a known adipocyte marker and then  
 CC performing a yeast 2-hybrid selection to isolate prey proteins encoded by  
 CC members of an adipocyte cDNA library. The proteins are designated SID  
 CC (RTM) (selected interacting domains) proteins. Also included are a  
 CC polynucleotide encoding a polypeptide in the adipocyte cells, a  
 CC recombinant host cell expressing at least one of the interacting  
 CC polypeptides of the complex, selecting a modulating compound in adipocyte  
 CC cells, a SID (RTM) polypeptide comprising any of the 738 amino acid  
 CC sequences given in the specification (including its fragment or variant),  
 CC a SID (RTM) polynucleotide comprising any of the 738 nucleotide sequences  
 CC given in the specification (including its fragment or variant), a vector  
 CC comprising the SID (RTM) polynucleotide, a recombinant host cell  
 CC comprising the vector, a protein chip comprising the polypeptides and a  
 CC record comprising all or part of the data, listed in the specification.  
 CC The complex, polypeptides, polynucleotides and compounds are useful for  
 CC preventing or treating metabolic disorders such as obesity or diabetes.  
 CC The polynucleotides are useful as probes or primers. The complex is  
 CC particularly useful for identifying selected interacting domains (SID  
 CC (RTM)) for screening drugs that modulate the protein interaction, thus  
 CC exhibiting the therapeutic effect. The present sequence represents a SID  
 CC (prey) protein of the invention  
 XX  
 SQ Sequence 293 AA;

Query Match 100.0%; Score 48; DB 6; Length 293;  
 Best Local Similarity 100.0%; Pred. No. 2.9;



```
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 IGGERAGGF 9
DB 139 IGGERAGGF 147
|||||
RESULT 8
AAU23675
ID AAU23675 standard; protein; 309 AA.
XX AC
XX AAU23675;
XX 18-DEC-2001 (first entry)
XX
XX Novel human enzyme polypeptide #761.
XX
XX Human; oxidoreductase enzyme; transferase; hydrolase; lyase; isomerase;
XX ligase; hyperproliferative disorder; immunodeficiency disorder;
XX autoimmune disorder; neurological disorder; metabolic disorder;
XX inflammatory disorder; cardiovascular disorder; reproductive disorder;
XX blood-related disorder; infectious disorder; cytostatic; anti arthritic;
XX nephrotropic; anticoagulant.
XX
XX Homo sapiens.
XX
XX WO200155301-A2.
XX
XX 02-AUG-2001.
XX
XX 17-JAN-2001; 2001WO-US001239.
XX
XX 31-JAN-2000; 2000US-0179065P.
XX 04-FEB-2000; 2000US-0180628P.
XX 24-FEB-2000; 2000US-0184664P.
XX 02-MAR-2000; 2000US-0186350P.
XX 16-MAR-2000; 2000US-0189874P.
XX 17-MAR-2000; 2000US-0190076P.
XX 18-APR-2000; 2000US-01981123P.
XX 19-MAY-2000; 2000US-0205515P.
XX 07-JUN-2000; 2000US-0209467P.
XX 28-JUN-2000; 2000US-0214886P.
XX 30-JUN-2000; 2000US-0215135P.
XX 07-JUL-2000; 2000US-0216647P.
XX 07-JUL-2000; 2000US-0216880P.
XX 11-JUL-2000; 2000US-0217487P.
XX 11-JUL-2000; 2000US-0217496P.
XX 14-JUL-2000; 2000US-0220963P.
XX 26-JUL-2000; 2000US-0220963P.
XX 26-JUL-2000; 2000US-0220964P.
XX 14-AUG-2000; 2000US-0224518P.
XX 14-AUG-2000; 2000US-0224519P.
XX 14-AUG-2000; 2000US-0225213P.
XX 14-AUG-2000; 2000US-0225214P.
XX 14-AUG-2000; 2000US-0225266P.
XX 14-AUG-2000; 2000US-0225267P.
XX 14-AUG-2000; 2000US-0225288P.
XX 14-AUG-2000; 2000US-0225270P.
XX 14-AUG-2000; 2000US-0225447P.
XX 14-AUG-2000; 2000US-0225757P.
XX 14-AUG-2000; 2000US-0225758P.
XX 14-AUG-2000; 2000US-0225759P.
XX 18-AUG-2000; 2000US-0226279P.
XX 22-AUG-2000; 2000US-0226681P.
XX 22-AUG-2000; 2000US-0226868P.
XX 22-AUG-2000; 2000US-0227182P.
XX 23-AUG-2000; 2000US-0227009P.
XX 30-AUG-2000; 2000US-0228924P.
XX 01-SEP-2000; 2000US-0229287P.
XX 01-SEP-2000; 2000US-0229343P.
XX 01-SEP-2000; 2000US-0229344P.
XX 01-SEP-2000; 2000US-0229345P.
XX 05-SEP-2000; 2000US-0229509P.
XX 05-SEP-2000; 2000US-0229513P.
XX 06-SEP-2000; 2000US-0230437P.
XX 06-SEP-2000; 2000US-0230438P.
XX 08-SEP-2000; 2000US-0231242P.
XX 08-SEP-2000; 2000US-0231243P.
XX 08-SEP-2000; 2000US-0231244P.
XX 08-SEP-2000; 2000US-0231413P.
XX 08-SEP-2000; 2000US-0231414P.
XX 08-SEP-2000; 2000US-0232080P.
XX 08-SEP-2000; 2000US-0232081P.
XX 12-SEP-2000; 2000US-0231968P.
XX 14-SEP-2000; 2000US-0232397P.
XX 14-SEP-2000; 2000US-0232398P.
XX 14-SEP-2000; 2000US-0232399P.
XX 14-SEP-2000; 2000US-0232400P.
XX 14-SEP-2000; 2000US-0232401P.
XX 14-SEP-2000; 2000US-0233063P.
XX 14-SEP-2000; 2000US-0233064P.
XX 14-SEP-2000; 2000US-0233065P.
XX 21-SEP-2000; 2000US-0234223P.
XX 21-SEP-2000; 2000US-0234274P.
XX 25-SEP-2000; 2000US-0234997P.
XX 25-SEP-2000; 2000US-0234998P.
XX 26-SEP-2000; 2000US-0235484P.
XX 27-SEP-2000; 2000US-0235834P.
XX 27-SEP-2000; 2000US-0235836P.
XX 29-SEP-2000; 2000US-0236327P.
XX 29-SEP-2000; 2000US-0236367P.
XX 29-SEP-2000; 2000US-0236368P.
XX 29-SEP-2000; 2000US-0236369P.
XX 29-SEP-2000; 2000US-0236370P.
XX 02-OCT-2000; 2000US-0236802P.
XX 02-OCT-2000; 2000US-0237037P.
XX 02-OCT-2000; 2000US-0237038P.
XX 02-OCT-2000; 2000US-0237039P.
XX 12-OCT-2000; 2000US-0237040P.
XX 13-OCT-2000; 2000US-0239935P.
XX 13-OCT-2000; 2000US-0239937P.
XX 20-OCT-2000; 2000US-0240960P.
XX 20-OCT-2000; 2000US-0241221P.
XX 20-OCT-2000; 2000US-0241785P.
XX 20-OCT-2000; 2000US-0241786P.
XX 20-OCT-2000; 2000US-0241808P.
XX 20-OCT-2000; 2000US-0241809P.
XX 01-NOV-2000; 2000US-0244617P.
XX 08-NOV-2000; 2000US-0246474P.
XX 08-NOV-2000; 2000US-0246475P.
XX 08-NOV-2000; 2000US-0246476P.
XX 08-NOV-2000; 2000US-0246477P.
XX 08-NOV-2000; 2000US-0246478P.
XX 08-NOV-2000; 2000US-0246523P.
XX 08-NOV-2000; 2000US-0246524P.
XX 08-NOV-2000; 2000US-0246525P.
XX 08-NOV-2000; 2000US-0246526P.
XX 08-NOV-2000; 2000US-0246527P.
XX 08-NOV-2000; 2000US-0246528P.
XX 08-NOV-2000; 2000US-0246532P.
XX 08-NOV-2000; 2000US-0246609P.
XX 08-NOV-2000; 2000US-0246610P.
XX 08-NOV-2000; 2000US-0246611P.
XX 17-NOV-2000; 2000US-0246613P.
XX 17-NOV-2000; 2000US-0249207P.
XX 17-NOV-2000; 2000US-0249208P.
XX 17-NOV-2000; 2000US-0249211P.
XX 17-NOV-2000; 2000US-0249212P.
XX 17-NOV-2000; 2000US-0249213P.
XX 17-NOV-2000; 2000US-0249214P.
XX 17-NOV-2000; 2000US-0249215P.
XX 17-NOV-2000; 2000US-0249216P.
```

PR 17-NOV-2000; 2000US-0249217P.  
 PR 17-NOV-2000; 2000US-0249218P.  
 PR 17-NOV-2000; 2000US-0249244P.  
 PR 17-NOV-2000; 2000US-0249245P.  
 PR 17-NOV-2000; 2000US-0249264P.  
 PR 17-NOV-2000; 2000US-0249265P.  
 PR 17-NOV-2000; 2000US-0249297P.  
 PR 17-NOV-2000; 2000US-0249299P.  
 PR 17-NOV-2000; 2000US-0249300P.  
 PR 01-DEC-2000; 2000US-0250160P.  
 PR 05-DEC-2000; 2000US-0251030P.  
 PR 05-DEC-2000; 2000US-0251388P.  
 PR 05-DEC-2000; 2000US-0251988P.  
 PR 06-DEC-2000; 2000US-0256719P.  
 PR 06-DEC-2000; 2000US-0251479P.  
 PR 08-DEC-2000; 2000US-0251856P.  
 PR 08-DEC-2000; 2000US-0251868P.  
 PR 08-DEC-2000; 2000US-0251869P.  
 PR 08-DEC-2000; 2000US-0251989P.  
 PR 08-DEC-2000; 2000US-0251990P.  
 PR 11-DEC-2000; 2000US-0254097P.  
 PR 05-JAN-2001; 2001US-0259678P.  
 XX  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 PI Rosen CA, Barash SC, Ruben SM;  
 XX  
 XX WPI; 2001-465566/50.  
 DR N-PSDB; AAS41545.  
 XX  
 XX Novel polypeptides and polynucleotides useful for diagnosing, preventing,  
 PT treating neural, immune system, muscular, reproductive, pulmonary,  
 PT cardiovascular, renal, proliferative disorders and cancerous diseases.  
 XX  
 PS Claim 11; SEQ ID NO 1671; 1180pp; English.  
 XX  
 CC The present invention relates to the isolation of novel human enzyme  
 CC polypeptides, and the cDNA (AAS40785-AAS41684) and genomic sequences  
 CC encoding them. The enzyme polypeptides of the invention may comprise the  
 CC functional classes of oxidoreductases, transferases, hydrolases, lyases,  
 CC isomerases or ligases. The sequences of the invention are useful in the  
 CC diagnosis, treatment, prevention and/or prognosis of a wide range of  
 CC disorders including hyperproliferative disorders (e.g. cancer),  
 CC immunodeficiency disorders (e.g. AIDS) autoimmune disorders (e.g.  
 CC arthritis), neurological disorders (e.g. Alzheimer's disease), metabolic  
 CC disorders (e.g. phenylketonuria), inflammatory disorders (e.g. asthma),  
 CC cardiovascular disorders (e.g. atherosclerosis), blood-related disorders  
 CC (e.g. haemophilia), reproductive disorders (e.g. infertility) and  
 CC infectious disorders (e.g. Influenza). The polynucleotides of the  
 CC invention can also be used in gene therapy. AAU22915-AAU23814 represent  
 CC the novel human enzyme polypeptides of the invention. Note: The sequence  
 CC data for this patent did not form part of the printed specification, but  
 CC was obtained in electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 SQ Sequence 309 AA;  
 Query Match 100.0%; Score 48; DB 4; Length 309;  
 Best Local Similarity 100.0%; Pred. No. 3.1;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 IGGEKAGGF 9  
 Db 51 IGGEKAGGF 59  
 RESULT 9  
 ABG60248  
 ID ABG60248 standard; protein; 309 AA.  
 XX  
 AC ABG60248;  
 XX  
 DT 13-AUG-2002 (first entry)

XX  
 DE Human ovarian antigen #10.  
 XX  
 KW Human; ovarian antigen; ovary disorder; breast disorder;  
 KW neoplastic disorder; cancer; infectious disease; inflammatory disease;  
 KW reproductive system disorder; autoimmune disorder; Alzheimer's disease;  
 KW blood-related disorder; hyperproliferative disorder; hair loss;  
 KW urinary system disorder; cardiovascular disorder; arrhythmia;  
 KW respiratory disorder; musculoskeletal system disorder;  
 KW neural activity disorder; neurological disorder; endocrine disorder;  
 KW gastrointestinal disorder; liver disorder; pancreatic disorder;  
 KW gall bladder disorder; large intestine disorder; developmental disorder;  
 KW inherited disorder; wound healing; skin aging; food additive;  
 KW preservative.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200155329-A2.  
 XX  
 XX 02-AUG-2001.  
 XX  
 XX 17-JAN-2001; 2001WO-US001360.  
 XX  
 PR 31-JAN-2000; 2000US-0179065P.  
 PR 04-FEB-2000; 2000US-0180628P.  
 PR 07-JUN-2000; 2000US-0209467P.  
 PR 14-SEP-2000; 2000US-0232388P.  
 PR 17-NOV-2000; 2000US-0249300P.  
 PR 01-DEC-2000; 2000US-0250160P.  
 PR 08-DEC-2000; 2000US-0251868P.  
 PR 08-DEC-2000; 2000US-0251990P.  
 XX  
 XX (HUMA-) HUMAN GENOME SCI INC.  
 PA  
 XX Rosen CA, Barash SC, Ruben SM;  
 PI  
 PI WPI; 2001-476195/51.  
 DR N-PSDB; ABK72051.  
 XX  
 XX Novel isolated human ovarian related polypeptide useful for  
 PT diagnosis/treatment of disorders of ovary and breast such as neoplastic  
 PT disorders, infectious diseases, inflammatory diseases, and reproductive  
 PT disorders.  
 XX  
 PS Claim 11; SEQ ID NO 78; 524pp; English.  
 XX  
 CC The invention relates to isolated ovarian related polypeptide (ovarian  
 CC antigen) comprising a sequence at least 90% identical to a sequence  
 CC selected from a polypeptide fragment, domain, epitope or full length  
 CC protein of a sequence (S1) appearing as ABG60239-ABG60296 having  
 CC biological activity, or a variant, allelic variant or species homologue  
 CC of S1. Also included are the cDNA clones encoding the proteins of S1. S1,  
 CC an anti-S1 antibody and the cDNA are useful for diagnosing, preventing,  
 CC treating or ameliorating a medical condition in mammalian subject  
 CC especially diseases and/or disorders of the ovary and/or breast such as  
 CC neoplastic diseases (such as ovarian Krukenberg tumour and cancer),  
 CC infectious diseases (e.g. mastitis, oophoritis), inflammatory diseases  
 CC (e.g., abscesses), reproductive system disorders (Paget's disease),  
 CC autoimmune disorders (systemic lupus erythematosus, rheumatoid  
 CC arthritis), blood-related disorders (sickle cell anaemia),  
 CC hyperproliferative disorders, urinary system disorders  
 CC (glomerulonephritis), cardiovascular disorders (arrhythmias), respiratory  
 CC disorders, musculoskeletal system disorders (arthritis), and  
 CC neurological disorders (Alzheimer's disease and Parkinson's disease),  
 CC endocrine disorders (Addison's disease), gastrointestinal disorders  
 CC (inflammatory disorders), liver disorders (biliary liver cirrhosis),  
 CC pancreatic and gall bladder disorders, diseases of the large intestine,  
 CC developmental and inherited disorders, diseases at the cellular level,  
 CC and wound healing and epithelial cell proliferation. They are also useful  
 CC to prevent skin aging, for preventing hair loss, to maintain organs  
 CC before transplantation or for supporting cell culture of primary tissues,  
 CC to modulate mammalian characteristics such as body height, to modulate  
 CC mammalian metabolism, to change a mammal's mental or physical state, and

CC as food additive or preservative. The present sequence represents an  
 CC ovarian antigen, SI protein of the invention. Note: The sequence data for  
 CC this patent did not form part of the printed specification, but was  
 CC obtained in electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences. Note: The sequence data for  
 CC this patent did not form part of the printed specification, but was  
 CC obtained in electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences  
 XX Sequence 309 AA;  
 SQ

Query Match 100.0%; Score 48; DB 4; Length 309;  
 Best Local Similarity 100.0%; Pred. No. 3.1;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IGGKAGGF 9  
 Db 51 IGGKAGGF 59  
 |||||

RESULT 10  
 ABG61719  
 ID ABG61719 standard; protein; 309 AA.  
 XX AC  
 AC ABG61719;  
 XX DT  
 DT 26-AUG-2002 (first entry)  
 XX DE  
 DE Novel ovarian related polypeptide #10.  
 XX KW  
 KW Ovarian related polypeptide; neoplastic disorder; tumour; ovarian cancer;  
 KW hyperproliferative disorder; adult acute lymphocytic leukaemia;  
 KW breast cancer; reproductive system disorder; tuberculosis; arthritis;  
 KW immune system disorder; Chediak-Higashi's syndrome; neonatal neutropenia;  
 KW autoimmune disorder; Hashimoto's thyroiditis; inflammatory disorder;  
 KW septic shock; multiple sclerosis; central nervous system disorder;  
 KW neurological disorder; allergy; Parkinson's disease; Alzheimer's disease;  
 KW cardiovascular disorder; atherosclerosis; blood related disorder;  
 KW respiratory disorder; urinary system disorder; musculoskeletal disorder;  
 KW osteoporosis; wound healing; endocrine disorder; infectious disease;  
 KW gastrointestinal disorder; transplantation; food additive; preservative.  
 XX OS  
 OS Homo sapiens.  
 XX PN  
 PN US2002045230-A1.  
 XX PD  
 PD 18-APR-2002.  
 XX PF  
 PF 20-JUL-2001; 2001US-00908711.  
 XX PR  
 PR 31-JAN-2000; 2000US-0179065P.  
 PR 04-FEB-2000; 2000US-0180628P.  
 PR 24-FEB-2000; 2000US-0184664P.  
 PR 02-MAR-2000; 2000US-0186350P.  
 PR 16-MAR-2000; 2000US-0189874P.  
 PR 17-MAR-2000; 2000US-0190076P.  
 PR 18-APR-2000; 2000US-0198123P.  
 PR 19-MAY-2000; 2000US-0205515P.  
 PR 07-JUN-2000; 2000US-0209467P.  
 PR 28-JUN-2000; 2000US-0214886P.  
 PR 30-JUN-2000; 2000US-0215135P.  
 PR 07-JUL-2000; 2000US-0216647P.  
 PR 07-JUL-2000; 2000US-0216880P.  
 PR 11-JUL-2000; 2000US-0217487P.  
 PR 11-JUL-2000; 2000US-0217496P.  
 PR 14-JUL-2000; 2000US-0218290P.  
 PR 26-JUL-2000; 2000US-0220963P.  
 PR 26-JUL-2000; 2000US-0220964P.  
 PR 14-AUG-2000; 2000US-0224518P.  
 PR 14-AUG-2000; 2000US-0224519P.  
 PR 14-AUG-2000; 2000US-0225213P.  
 PR 14-AUG-2000; 2000US-0225214P.  
 PR 14-AUG-2000; 2000US-0225266P.

PR 14-AUG-2000; 2000US-0225267P.  
 PR 14-AUG-2000; 2000US-0225268P.  
 PR 14-AUG-2000; 2000US-0225270P.  
 PR 14-AUG-2000; 2000US-0225447P.  
 PR 14-AUG-2000; 2000US-0225757P.  
 PR 14-AUG-2000; 2000US-0225758P.  
 PR 14-AUG-2000; 2000US-0225759P.  
 PR 18-AUG-2000; 2000US-0226279P.  
 PR 22-AUG-2000; 2000US-0226681P.  
 PR 22-AUG-2000; 2000US-0226686P.  
 PR 22-AUG-2000; 2000US-0227182P.  
 PR 23-AUG-2000; 2000US-0227009P.  
 PR 30-AUG-2000; 2000US-0228924P.  
 PR 01-SEP-2000; 2000US-0229287P.  
 PR 01-SEP-2000; 2000US-0229343P.  
 PR 01-SEP-2000; 2000US-0229344P.  
 PR 01-SEP-2000; 2000US-0229345P.  
 PR 05-SEP-2000; 2000US-0229509P.  
 PR 05-SEP-2000; 2000US-0229513P.  
 PR 06-SEP-2000; 2000US-0230437P.  
 PR 06-SEP-2000; 2000US-0230438P.  
 PR 08-SEP-2000; 2000US-0231242P.  
 PR 08-SEP-2000; 2000US-0231243P.  
 PR 08-SEP-2000; 2000US-0231244P.  
 PR 08-SEP-2000; 2000US-0231413P.  
 PR 08-SEP-2000; 2000US-0231414P.  
 PR 08-SEP-2000; 2000US-0232080P.  
 PR 08-SEP-2000; 2000US-0232081P.  
 PR 12-SEP-2000; 2000US-0231968P.  
 PR 14-SEP-2000; 2000US-0232397P.  
 PR 14-SEP-2000; 2000US-0232398P.  
 PR 14-SEP-2000; 2000US-0232399P.  
 PR 14-SEP-2000; 2000US-0232400P.  
 PR 14-SEP-2000; 2000US-0232401P.  
 PR 14-SEP-2000; 2000US-0233063P.  
 PR 14-SEP-2000; 2000US-0233064P.  
 PR 21-SEP-2000; 2000US-0233065P.  
 PR 21-SEP-2000; 2000US-0234223P.  
 PR 21-SEP-2000; 2000US-0234274P.  
 PR 25-SEP-2000; 2000US-0234997P.  
 PR 25-SEP-2000; 2000US-0234998P.  
 PR 26-SEP-2000; 2000US-0235484P.  
 PR 27-SEP-2000; 2000US-0235834P.  
 PR 27-SEP-2000; 2000US-0235836P.  
 PR 29-SEP-2000; 2000US-0236327P.  
 PR 29-SEP-2000; 2000US-0236367P.  
 PR 29-SEP-2000; 2000US-0236368P.  
 PR 29-SEP-2000; 2000US-0236369P.  
 PR 29-SEP-2000; 2000US-0236370P.  
 PR 02-OCT-2000; 2000US-0236802P.  
 PR 02-OCT-2000; 2000US-0237037P.  
 PR 02-OCT-2000; 2000US-0237038P.  
 PR 02-OCT-2000; 2000US-0237039P.  
 PR 02-OCT-2000; 2000US-0237040P.  
 PR 13-OCT-2000; 2000US-0239355P.  
 PR 13-OCT-2000; 2000US-0239937P.  
 PR 20-OCT-2000; 2000US-0240960P.  
 PR 20-OCT-2000; 2000US-0241221P.  
 PR 20-OCT-2000; 2000US-0241785P.  
 PR 20-OCT-2000; 2000US-0241786P.  
 PR 20-OCT-2000; 2000US-0241787P.  
 PR 20-OCT-2000; 2000US-0241808P.  
 PR 20-OCT-2000; 2000US-0241809P.  
 PR 01-NOV-2000; 2000US-0244617P.  
 PR 08-NOV-2000; 2000US-0246474P.  
 PR 08-NOV-2000; 2000US-0246475P.  
 PR 08-NOV-2000; 2000US-0246476P.  
 PR 08-NOV-2000; 2000US-0246477P.  
 PR 08-NOV-2000; 2000US-0246478P.  
 PR 08-NOV-2000; 2000US-0246523P.  
 PR 08-NOV-2000; 2000US-0246524P.  
 PR 08-NOV-2000; 2000US-0246525P.



CC (prey) protein of the invention  
 XX Sequence 339 AA;  
 SQ

Query Match 100.0%; Score 48; DB 6; Length 339;  
 Best Local Similarity 100.0%; Pred. No. 3.4;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IGGEKAGGF 9  
 |||||  
 DB 81 IGGEKAGGF 89

## RESULT 12

AAW12843  
 ID AAW12843 standard; peptide; 623 AA.

XX AC AAW12843;

DT 15-DEC-1997 (first entry)

XX DE Pro-alpha(III): (I) CP chimeric protein.

XX C-propeptide; recognition sequence; procollagen; monomer chain; therapy;  
 KW trimerised pro-alpha chain; fibril; procollagen suicide; wound healing;  
 KW fibrotic disease; human; chimeric protein.

OS Homo sapiens.

XX WO9708311-A1.

XX PD 06-MAR-1997.

XX DF 30-AUG-1996; 96WO-GB002122.

XX PR 31-AUG-1995; 95GB-00017773.

XX PR 23-MAR-1996; 96GB-00006152.

XX PR 14-JUN-1996; 96GB-00012476.

XX PA (UYMA-) UNIV VICTORIA MANCHESTER.

XX PI Bulleid N, Kadler K;

XX WPI; 1997-179268/16.

PT Novel pro:collagen mol. - comprising pro:collagen C-pro:peptide attached  
 PT to an alien collagen alpha-chain or non-collagen material, useful e.g.  
 PT for wound healing.

PS Example 1; Page 32-35; 69pp; English.

XX This sequence represents a chimeric procollagen molecule of the  
 CC invention. This sequence has the procollagen C-propeptide from the pro  
 CC -alpha2(I) chain attached to the pro-alpha(III) chain sequence. The C-  
 CC propeptide is implicated in the assembly of the monomer chains into  
 CC trimerised pro-alpha chains prior to cleavage of the N- and C-propeptides  
 CC and formation of collagen in fibril-forming pro-alpha chains. The C-  
 CC propeptides determine the type-specific assembly of the moieties to which  
 CC they are attached. The molecule of the invention comprises a first moiety  
 CC having procollagen C-propeptide activity attached to a second moiety,  
 CC which is an alien collagen alpha-chain or a non-collagen material. The  
 CC novel collagen molecule can be used for treatment or diagnosis in humans  
 CC or animals, especially for the treatment of procollagen suicide, as an  
 CC adhesive or implant, to promote (chronic) wound healing or fibrotic  
 CC diseases with reduced scarring or for use in photography, brewing,  
 CC foodstuffs or textiles. The novel collagen molecules, especially when  
 CC containing substitutions in the recognition site, may have significantly  
 CC altered properties and characteristics, such as different binding  
 CC kinetics or alpha-chain selection properties

XX SQ Sequence 623 AA;

Query Match 100.0%; Score 48; DB 2; Length 623;

Best Local Similarity 100.0%; Pred. No. 6.4;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IGGEKAGGF 9  
 |||||  
 DB 368 IGGEKAGGF 376

## RESULT 13

AAW12842  
 ID AAW12842 standard; protein; 626 AA.

XX AC AAW12842;

XX DT 15-DEC-1997 (first entry)

XX DE Truncated pro-alpha(III) chain.

XX C-propeptide; recognition sequence; procollagen; monomer chain; therapy;  
 KW trimerised pro-alpha chain; fibril; procollagen suicide; wound healing;  
 KW fibrotic disease; human.

XX OS Homo sapiens.

XX PN WO9708311-A1.

XX PD 06-MAR-1997.

XX PF 30-AUG-1996; 96WO-GB002122.

XX PR 31-AUG-1995; 95GB-00017773.

XX PR 23-MAR-1996; 96GB-00006152.

XX PR 14-JUN-1996; 96GB-00012476.

XX PA (UYMA-) UNIV VICTORIA MANCHESTER.

XX PI Bulleid N, Kadler K;

XX WPI; 1997-179268/16.

XX DR N-PSDB; AAT59892.

PT Novel pro:collagen mol. - comprising pro:collagen C-pro:peptide attached  
 PT to an alien collagen alpha-chain or non-collagen material, useful e.g.  
 PT for wound healing.

PS Example 1; Page 28-31; 69pp; English.

XX This sequence represents a truncated procollagen pro-alpha(III) chain  
 CC that can be used in the procollagen molecules of the invention. The C-  
 CC propeptide is implicated in the assembly of the monomer chains into  
 CC trimerised pro-alpha chains prior to cleavage of the N- and C-propeptides  
 CC and formation of collagen in fibril-forming pro-alpha chains. The C-  
 CC propeptides determine the type-specific assembly of the moieties to which  
 CC they are attached. The molecule of the invention comprises a first moiety  
 CC having procollagen C-propeptide activity attached to a second moiety,  
 CC which is an alien collagen alpha-chain or a non-collagen material. The  
 CC novel collagen molecule can be used for treatment or diagnosis in humans  
 CC or animals, especially for the treatment of procollagen suicide, as an  
 CC adhesive or implant, to promote (chronic) wound healing or fibrotic  
 CC diseases with reduced scarring or for use in photography, brewing,  
 CC foodstuffs or textiles. The novel collagen molecules, especially when  
 CC containing substitutions in the recognition site, may have significantly  
 CC altered properties and characteristics, such as different binding  
 CC kinetics or alpha-chain selection properties

XX SQ Sequence 626 AA;

Query Match 100.0%; Score 48; DB 2; Length 626;  
 Best Local Similarity 100.0%; Pred. No. 6.4;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IGGEKAGGF 9  
 |||||

DB	368	IGGERKAGGF 376	
RESULT 14			
ADE08475		ID ADE08475 standard; protein; 842 AA.	
XX	AC	ADE08475;	
XX	DT	29-JAN-2004 (first entry)	
XX	DE	Novel protein (useful for identifying genetic disorders) #630.	
XX	KW	novel gene; novel protein; tissue marker; molecular weight marker;	
XX	KW	chromosome marker; genetic disorder.	
XX	OS	Unidentified.	
XX	PN	WO2003054152-A2.	
XX	PD	03-JUL-2003.	
XX	PF	10-DEC-2002; 2002WO-US039555.	
XX	PR	10-DEC-2001; 2001US-0339739P.	
XX	PR	11-DEC-2001; 2001US-0339453P.	
XX	PR	14-MAR-2002; 2002US-0365091P.	
XX	PR	14-MAR-2002; 2002US-0365384P.	
XX	PR	12-APR-2002; 2002US-0372381P.	
XX	PR	12-APR-2002; 2002US-0372615P.	
XX	PR	22-APR-2002; 2002US-00128558.	
XX	PR	24-APR-2002; 2002US-0376045P.	
XX	PA	(HYSE-) HYSEQ INC.	
XX	PI	Tang YT, Asundi V, Goodrich RW, Ren F, Zhang J, Zhao QA, Wang J;	
XX	PI	Ghosh M, Xue AJ, Wehrman T, Weng G, Zhou P, Drmanac RT, Wang Z;	
XX	PI	Ma Y, Wang D, Chen R, Xu C, Boyle BJ;	
XX	DR	WPI; 2003-569235/53.	
XX	DR	N-PSDB; ADE07564.	
XX	PT	New polynucleotides, useful for expressing recombinant proteins for	
XX	PT	analysis, characterization or therapeutic use, or as markers for tissues	
XX	PT	in which the corresponding protein is preferentially expressed.	
XX	PS	Claim 20; SEQ ID NO 1541; 1177pp; English.	
XX	CC	The invention comprises the amino acid and coding sequences of novel	
XX	CC	proteins. The DNA and protein sequences of the invention are useful as;	
XX	CC	markers for tissues in which the corresponding protein is preferentially	
XX	CC	expressed; as molecular weight markers on gels; as chromosome markers or	
XX	CC	tags; to identify chromosomes or to map related gene positions; and to	
XX	CC	compare with endogenous DNA sequences in patients to identify potential	
XX	CC	genetic disorders. The present amino acid sequence represents a protein	
XX	CC	of the invention.	
XX	CC	Sequence 842 AA;	
XX	CC	Query Match 100.0%; Score 48; DB 7; Length 842;	
XX	CC	Best Local Similarity 100.0%; Pred. No. 8.7;	
XX	CC	Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
QY	1	IGGERKAGGF 9	
DB	584	IGGERKAGGF 592	
RESULT 15			
ABB50291		ID ABB50291 standard; protein; 1466 AA.	
XX	AC	ABB50291;	

QY 1 IGGERAGGF 9  
| | | | |  
Db 1208 IGGERAGGF 1216

Search completed: September 18, 2004, 04:25:32  
Job time : 49.1765 secs





GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 18, 2004, 04:20:54 ; Search time 14.8235 Seconds  
(without alignments)  
31.344 Million cell updates/sec

Title: US-10-615-959-41  
Perfect score: 48  
Sequence: 1 IGGEKAGGF 9

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA:\*

1: /cgn2\_6/ptodata/2/iaa/5A-COMB.pep.\*  
2: /cgn2\_6/ptodata/2/iaa/5B-COMB.pep.\*  
3: /cgn2\_6/ptodata/2/iaa/6A-COMB.pep.\*  
4: /cgn2\_6/ptodata/2/iaa/6B-COMB.pep.\*  
5: /cgn2\_6/ptodata/2/iaa/PCTUS-COMB.pep.\*  
6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	48	100.0	9	US-10-009-999A-41	Sequence 41, Appl
2	48	100.0	12	US-10-009-999A-43	Sequence 43, Appl
3	48	100.0	15	US-08-278-774-20	Sequence 20, Appl
4	48	100.0	623	US-09-029-348-3	Sequence 3, Appl
5	48	100.0	626	US-09-029-348-2	Sequence 2, Appl
6	42	87.5	8	US-10-009-999A-40	Sequence 40, Appl
7	42	87.5	11	US-10-009-999A-42	Sequence 42, Appl
8	42	87.5	11	US-10-009-999A-43	Sequence 43, Appl
9	42	87.5	1078	US-08-963-825-21	Sequence 21, Appl
10	42	87.5	1078	US-09-500-811-21	Sequence 21, Appl
11	42	87.5	1078	US-09-570-573-21	Sequence 21, Appl
12	42	87.5	1078	US-09-548-608-21	Sequence 21, Appl
13	38	79.2	674	US-09-193-637A-51	Sequence 51, Appl
14	38	79.2	674	US-09-252-991A-26476	Sequence 51, Appl
15	37	77.1	725	US-09-252-991A-27825	Sequence 26476, A
16	36	75.0	1504	US-09-252-991A-26608	Sequence 26476, A
17	35	72.9	10	US-08-702-105A-19	Sequence 26608, A
18	35	72.9	10	US-08-702-110A-19	Sequence 19, Appl
19	35	72.9	10	US-09-325-571-19	Sequence 19, Appl
20	35	72.9	10	US-09-848-585-19	Sequence 19, Appl
21	35	72.9	219	US-08-164-614A-9	Sequence 19, Appl
22	35	72.9	219	US-08-456-489B-9	Sequence 9, Appl
23	35	72.9	379	US-08-164-614A-8	Sequence 9, Appl
24	35	72.9	379	US-08-456-489B-8	Sequence 8, Appl
25	35	72.9	468	US-08-164-614A-7	Sequence 8, Appl
26	35	72.9	468	US-08-456-489B-7	Sequence 7, Appl
27	35	72.9	561	US-09-252-991A-28074	Sequence 28074, A

28 70.8 168 4 US-09-252-991A-20078 Sequence 20078, A  
29 70.8 284 4 US-09-252-991A-28522 Sequence 28522, A  
30 70.8 529 4 US-09-252-991A-27659 Sequence 27659, A  
31 70.8 540 4 US-09-252-991A-23300 Sequence 23300, A  
32 70.8 822 4 US-09-252-991A-22479 Sequence 22479, A  
33 70.8 70 4 US-09-621-976-6612 Sequence 22621, A  
34 68.8 221 4 US-09-252-991A-28805 Sequence 28805, A  
35 68.8 227 4 US-09-252-991A-30541 Sequence 30541, A  
36 68.8 254 4 US-09-252-991A-22621 Sequence 22621, A  
37 68.8 311 2 US-08-602-359A-41 Sequence 41, Appl  
38 68.8 551 4 US-09-252-991A-30116 Sequence 30116, A  
39 68.8 942 4 US-09-489-039A-13982 Sequence 13982, A  
40 66.7 6 4 US-10-009-999A-39 Sequence 39, Appl  
41 66.7 7 3 US-08-963-825-15 Sequence 15, Appl  
42 66.7 7 4 US-09-500-811-15 Sequence 15, Appl  
43 66.7 7 4 US-09-570-573-15 Sequence 15, Appl  
44 66.7 7 4 US-09-548-608-15 Sequence 15, Appl  
45 66.7 15 1 US-08-383-753-90 Sequence 90, Appl

## ALIGNMENTS

RESULT 1  
US-10-009-999A-41  
; Sequence 41, Application US/10009999A  
; Patent No. 6602980  
; GENERAL INFORMATION:  
; APPLICANT: Eyre, David R.  
; TITLE OF INVENTION: SYNTHETIC PEPTIDES FOR COLLAGEN RESORPTION ASSAYS  
; FILE REFERENCE: WROS-1-18220  
; CURRENT APPLICATION NUMBER: US/10/009,999A  
; PRIOR FILING DATE: 2003-01-22  
; PRIOR APPLICATION NUMBER: PCT/US99/29357  
; PRIOR FILING DATE: 1999-12-10  
; PRIOR APPLICATION NUMBER: US 09/335,098  
; PRIOR FILING DATE: 1999-06-17  
; PRIOR APPLICATION NUMBER: US 60/141,574  
; PRIOR FILING DATE: 1999-06-29  
; PRIOR APPLICATION NUMBER: US 60/142,274  
; PRIOR FILING DATE: 1999-07-02  
; PRIOR APPLICATION NUMBER: US 60/142,675  
; PRIOR FILING DATE: 1999-07-07  
; PRIOR APPLICATION NUMBER: US 09/385,740  
; PRIOR FILING DATE: 1999-08-30  
; NUMBER OF SEQ ID NOS: 45  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 41  
; LENGTH: 9  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: MISC FEATURE  
; OTHER INFORMATION: synthetic peptide corresponding to carboxy-terminal  
; OTHER INFORMATION: telopeptide sequence of human type III collagen  
US-10-009-999A-41

Query Match 100.0%; Score 48; DB 4; Length 9;  
Best Local Similarity 100.0%; Pred. No. 3e+05;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 IGGEKAGGF 9  
Db 1 IGGEKAGGF 9

RESULT 2  
US-10-009-999A-43  
; Sequence 43, Application US/10009999A  
; Patent No. 6602980  
; GENERAL INFORMATION:  
; APPLICANT: Eyre, David R.  
; TITLE OF INVENTION: SYNTHETIC PEPTIDES FOR COLLAGEN RESORPTION ASSAYS

```
FILE REFERENCE: WROS-1-18220
CURRENT APPLICATION NUMBER: US/10/009,999A
CURRENT FILING DATE: 2003-01-22
PRIOR APPLICATION NUMBER: PCT/US99/29357
PRIOR FILING DATE: 1999-12-10
PRIOR APPLICATION NUMBER: US 09/335,098
PRIOR FILING DATE: 1999-06-17
PRIOR APPLICATION NUMBER: US 60/141,574
PRIOR FILING DATE: 1999-06-29
PRIOR APPLICATION NUMBER: US 60/142,274
PRIOR FILING DATE: 1999-07-02
PRIOR APPLICATION NUMBER: US 60/142,675
PRIOR FILING DATE: 1999-07-07
PRIOR APPLICATION NUMBER: US 09/385,740
PRIOR FILING DATE: 1999-08-30
NUMBER OF SEQ ID NOS: 45
SOFTWARE: PatentIn version 3.2
SEQ ID NO 43
LENGTH: 12
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: MISC FEATURE
OTHER INFORMATION: Synthetic peptide corresponding to carboxy-terminal
OTHER INFORMATION: telopeptide sequence of human type III collagen
US-10-009-999A-43

Query Match          100.0%; Score 48; DB 4; Length 12;
Best Local Similarity 100.0%; Pred. No. 0.038;
Matches          9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 IGGEKAGGF 9
        |||||
DB      4 IGGEKAGGF 12

RESULT 3
US-08-278-774-20
Sequence 20, Application US/08278774
Patent No. 6653450
GENERAL INFORMATION:
APPLICANT: Berg, Richard A
APPLICANT: Toman, David P
APPLICANT: Wallace, Donald
TITLE OF INVENTION: MUTATED RECOMBINANT COLLAGENS
NUMBER OF SEQUENCES: 33
CORRESPONDENCE ADDRESS:
ADDRESSEE: COLLAGEN CORPORATION
STREET: 2500 Faber Place
CITY: Palo Alto
STATE: California
COUNTRY: USA
ZIP: 94303
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/278,774
FILING DATE: 22-JUL-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Rafayko, Kathi L
REGISTRATION NUMBER: 36,644
REFERENCE/DOCKET NUMBER: 94-018
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 354-4642
TELEFAX: (415) 354-4752
TELEX:
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 15 amino acids

TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-278-774-20

Query Match          100.0%; Score 48; DB 4; Length 15;
Best Local Similarity 100.0%; Pred. No. 0.047;
Matches          9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 IGGEKAGGF 9
        |||||
DB      2 IGGEKAGGF 10

RESULT 4
US-09-029-348-3
Sequence 3, Application US/09029348
Patent No. 6171827
GENERAL INFORMATION:
APPLICANT: THE VICTORIA UNIVERSITY OF MANCHESTER
TITLE OF INVENTION: NOVEL PROCOLLAGENS
FILE REFERENCE: 0087857PUS LISTING
CURRENT APPLICATION NUMBER: US/09/029,348
CURRENT FILING DATE: 1998-05-07
NUMBER OF SEQ ID NOS: 20
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 3
LENGTH: 623
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: SEQUENCE
OTHER INFORMATION: DERIVED FROM CDNA OF PROCOLLAGENS
US-09-029-348-3

Query Match          100.0%; Score 48; DB 3; Length 623;
Best Local Similarity 100.0%; Pred. No. 1.8;
Matches          9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 IGGEKAGGF 9
        |||||
DB      368 IGGEKAGGF 376

RESULT 5
US-09-029-348-2
Sequence 2, Application US/09029348
Patent No. 6171827
GENERAL INFORMATION:
APPLICANT: THE VICTORIA UNIVERSITY OF MANCHESTER
TITLE OF INVENTION: NOVEL PROCOLLAGENS
FILE REFERENCE: 0087857PUS LISTING
CURRENT APPLICATION NUMBER: US/09/029,348
CURRENT FILING DATE: 1998-05-07
NUMBER OF SEQ ID NOS: 20
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 2
LENGTH: 626
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: SEQUENCE
OTHER INFORMATION: DERIVED FROM CDNA OF PROCOLLAGENS
US-09-029-348-2

Query Match          100.0%; Score 48; DB 3; Length 626;
Best Local Similarity 100.0%; Pred. No. 1.8;
Matches          9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 IGGEKAGGF 9
        |||||
DB      368 IGGEKAGGF 376
```

```
RESULT 6
US-10-009-999A-40
; Sequence 40, Application US/10009999A
; Patent No. 6602980
; GENERAL INFORMATION:
; APPLICANT: Eyre, David R.
; TITLE OF INVENTION: SYNTHETIC PEPTIDES FOR COLLAGEN RESORPTION ASSAYS
; FILE REFERENCE: WROS-1-18220
; CURRENT APPLICATION NUMBER: US/10/009,999A
; CURRENT FILING DATE: 2003-01-22
; PRIOR APPLICATION NUMBER: PCT/US99/29357
; PRIOR FILING DATE: 1999-12-10
; PRIOR APPLICATION NUMBER: US 09/335,098
; PRIOR FILING DATE: 1999-06-17
; PRIOR APPLICATION NUMBER: US 60/141,574
; PRIOR FILING DATE: 1999-06-29
; PRIOR APPLICATION NUMBER: US 60/142,274
; PRIOR FILING DATE: 1999-07-02
; PRIOR APPLICATION NUMBER: US 60/142,675
; PRIOR FILING DATE: 1999-07-07
; PRIOR APPLICATION NUMBER: US 09/385,740
; PRIOR FILING DATE: 1999-08-30
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 40
; LENGTH: 8
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: MISC FEATURE
; OTHER INFORMATION: synthetic peptide corresponding to carboxy-terminal
; OTHER INFORMATION: telopeptide sequence of human type III collagen
US-10-009-999A-40

Query Match      87.5%; Score 42; DB 4; Length 8;
Best Local Similarity 100.0%; Pred. No. 3e+05;
Matches      8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 IGGEKAGG 8
      |||||
DB      1 IGGEKAGG 8

RESULT 7
US-10-009-999A-42
; Sequence 42, Application US/10009999A
; Patent No. 6602980
; GENERAL INFORMATION:
; APPLICANT: Eyre, David R.
; TITLE OF INVENTION: SYNTHETIC PEPTIDES FOR COLLAGEN RESORPTION ASSAYS
; FILE REFERENCE: WROS-1-18220
; CURRENT APPLICATION NUMBER: US/10/009,999A
; CURRENT FILING DATE: 2003-01-22
; PRIOR APPLICATION NUMBER: PCT/US99/29357
; PRIOR FILING DATE: 1999-12-10
; PRIOR APPLICATION NUMBER: US 09/335,098
; PRIOR FILING DATE: 1999-06-17
; PRIOR APPLICATION NUMBER: US 60/141,574
; PRIOR FILING DATE: 1999-06-29
; PRIOR APPLICATION NUMBER: US 60/142,274
; PRIOR FILING DATE: 1999-07-02
; PRIOR APPLICATION NUMBER: US 60/142,675
; PRIOR FILING DATE: 1999-07-07
; PRIOR APPLICATION NUMBER: US 09/385,740
; PRIOR FILING DATE: 1999-08-30
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 42
; LENGTH: 11
; TYPE: PRT
; ORGANISM: Homo sapiens
```

```
; FEATURE:
; NAME/KEY: MISC FEATURE
; OTHER INFORMATION: synthetic peptide corresponding to carboxy-terminal
; OTHER INFORMATION: telopeptide sequence of human type III collagen
US-10-009-999A-42

Query Match      87.5%; Score 42; DB 4; Length 11;
Best Local Similarity 100.0%; Pred. No. 0.35;
Matches      8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 IGGEKAGG 8
      |||||
DB      4 IGGEKAGG 11

RESULT 8
5473052-20
; Patent No. 5473052
; APPLICANT: EYRE, DAVID R.
; TITLE OF INVENTION: ANTIGEN-BINDING FRAGMENTS OF AN ANTIBODY
; TO TYPE-I COLLAGEN AMINO-TERMINAL TELEPEPTIDE
; NUMBER OF SEQUENCES: 30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/221,705
; FILING DATE: 01-APR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 614,719
; FILING DATE: 21-NOV-1990
; APPLICATION NUMBER: 444,881
; FILING DATE: 01-DEC-1989
; APPLICATION NUMBER: 118,234
; FILING DATE: 06-NOV-1987
; SEQ ID NO:20
; LENGTH: 11
5473052-20

Query Match      87.5%; Score 42; DB 6; Length 11;
Best Local Similarity 88.9%; Pred. No. 0.35;
Matches      8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 IGGEKAGGF 9
      |||||
DB      2 IGGEKAGGF 10

RESULT 9
US-08-963-825-21
; Sequence 21, Application US/08963825
; Patent No. 6110889
; GENERAL INFORMATION:
; APPLICANT: Qvist, Per
; APPLICANT: Bonde, Martin
; TITLE OF INVENTION: A Method for Assaying Collagen Fragments
; TITLE OF INVENTION: in Body Fluids, A Test Kit and Means for Carrying Out the
; TITLE OF INVENTION: Method and Use of the Method to Diagnose the Presence of
; TITLE OF INVENTION: Disorders Associated with the Metabolism of
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Darby & Darby PC
; STREET: 805 Third Avenue
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/963,825
; FILING DATE:
; CLASSIFICATION: 436
```

Query Match 87.5%; Score 42; DB 4; Length 1078;  
Best Local Similarity 88.9%; Pred. No. 31;  
Matches 8; Conservative 0; Mismatches 1; Indels

```
QY      1 IGGEKAGGF 9
Db      1061 IGAEKAGGF 1069

RESULT 12
US-09-548-608-21
; Sequence 21, Application US/09548608
; Patent No. 6355442
; GENERAL INFORMATION:
; APPLICANT: Qvist, Per
; APPLICANT: Bonde, Martin
; TITLE OF INVENTION: A Method for Assaying Collagen Fragments
; TITLE OF INVENTION: in Body Fluids, A Test Kit and Means for Carrying Out the
; TITLE OF INVENTION: Method and Use of the Method to Diagnose the Presence of
; TITLE OF INVENTION: Disorders Associated with the Metabolism of
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Darby & Darby PC
; STREET: 805 Third Avenue
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/548,608
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/187,319
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Gogoris, Adda C
; REGISTRATION NUMBER: 29,714
; REFERENCE/DOCKET NUMBER: 4305/08701
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-527-7700
; TELEFAX: 212-753-6237
; TELEX: 236687
; INFORMATION FOR SEQ ID NO: 21:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1078 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; IMMEDIATE SOURCE:
; CLONE: COLLAGEN ALPHA 1 (III)
US-09-548-608-21

Query Match      87.5%; Score 42; DB 4; Length 1078;
Best Local Similarity 88.9%; Pred. No. 31;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 IGGEKAGGF 9
Db      1061 IGAEKAGGF 1069

RESULT 13
US-09-199-637A-51
; Sequence 51, Application US/09199637A
; Patent No. 6355411
; GENERAL INFORMATION:
; APPLICANT: Ausubel, Frederick
; APPLICANT: Goodman, Howard M.
```

```
; APPLICANT: Rahme, Laurence G.
; APPLICANT: Mahajan-Miklos, Shalina
; APPLICANT: Tan, Man-Wah
; APPLICANT: Cao, Hui
; APPLICANT: Drenkard, Eliana
; APPLICANT: Tsongalis, John
; TITLE OF INVENTION: VIRULENCE-ASSOCIATED NUCLEIC ACID
; TITLE OF INVENTION: SEQUENCES AND USES THEREOF
; FILE REFERENCE: 00786/361002
; CURRENT APPLICATION NUMBER: US/09/199,637A
; CURRENT FILING DATE: 1998-11-25
; PRIOR APPLICATION NUMBER: 60/066,517
; PRIOR FILING DATE: 1997-11-25
; NUMBER OF SEQ ID NOS: 437
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 51
; LENGTH: 674
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-199-637A-51

Query Match      79.2%; Score 38; DB 4; Length 674;
Best Local Similarity 75.0%; Pred. No. 91;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY      1 IGGEKAGG 8
Db      413 VGGDKAGG 420

RESULT 14
US-09-252-991A-26476
; Sequence 26476, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 26476
; LENGTH: 674
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-26476

Query Match      79.2%; Score 38; DB 4; Length 674;
Best Local Similarity 75.0%; Pred. No. 91;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY      1 IGGEKAGG 8
Db      413 VGGDKAGG 420

RESULT 15
US-09-252-991A-27825
; Sequence 27825, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
```

Mon Sep 20 11:05:44 2004

```
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 27825
; LENGTH: 725
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-27825

Query Match      77.1%; Score 37; DB 4; Length 725;
Best Local Similarity 66.7%; Pred. No. 1.4e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY      1 IGGEKAGGF 9
      :|:|:|:|
Db      523 VGAEQAGGF 531

Search completed: September 18, 2004, 04:32:22
Job time : 15.8235 secs
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 18, 2004, 04:29:56 ; Search time 42.7059 Seconds  
(without alignments)  
67.677 Million cell updates/sec

Title: US-10-615-959-41  
Perfect score: 48  
Sequence: 1 IGGKAGGF 9

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1342398 seqs, 321133274 residues

Total number of hits satisfying chosen parameters: 1342398

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications AA:  
1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*  
2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*  
8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/2/pubpaa/US09A\_PUBCOMB.pep.\*  
10: /cgn2\_6/ptodata/2/pubpaa/US09B\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/2/pubpaa/US09C\_PUBCOMB.pep.\*  
12: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep.\*  
13: /cgn2\_6/ptodata/2/pubpaa/US10A\_PUBCOMB.pep.\*  
14: /cgn2\_6/ptodata/2/pubpaa/US10B\_PUBCOMB.pep.\*  
15: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pep.\*  
16: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep.\*  
17: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*  
18: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	48	100.0	9	12	US-10-615-959-41
2	48	100.0	12	12	US-10-615-959-43
3	48	100.0	25	15	US-10-366-125-11
4	48	100.0	262	15	US-10-366-125-19
5	48	100.0	309	9	US-09-908-711-78
6	48	100.0	1466	12	US-09-918-715-226
7	48	100.0	1466	12	US-10-257-021-72
8	48	100.0	1466	14	US-10-177-293-68
9	48	100.0	1466	14	US-10-301-822-33
10	48	100.0	1466	16	US-10-357-851-3
11	48	100.0	1466	16	US-10-358-024-3
12	48	100.0	1466	16	US-10-734-564-103
13	47	97.9	1466	15	US-10-402-089-12
14	47	97.9	1466	15	US-10-402-072A-12
15	42	87.5	8	12	US-10-615-959-40

16	42	87.5	11	12	US-10-615-959-42	Sequence 42, Appl
17	42	87.5	1078	14	US-10-058-124-21	Sequence 21, Appl
18	41	85.4	691	12	US-10-282-122A-47568	Sequence 47568, A
19	41	85.4	1466	15	US-10-402-089-4	Sequence 4, Appl
20	41	85.4	1466	15	US-10-402-089-6	Sequence 6, Appl
21	41	85.4	1466	15	US-10-402-072A-4	Sequence 4, Appl
22	41	85.4	1466	15	US-10-402-072A-6	Sequence 6, Appl
23	39	81.2	53	12	US-10-424-599-201930	Sequence 201930,
24	38	79.2	51	11	US-09-864-408A-5882	Sequence 5882, Ap
25	38	79.2	322	16	US-10-437-963-137654	Sequence 137654,
26	38	79.2	674	10	US-09-975-719-51	Sequence 51, Appl
27	38	79.2	696	16	US-10-437-963-189506	Sequence 189506,
28	37	77.1	225	16	US-10-437-963-106003	Sequence 106003,
29	37	77.1	488	16	US-10-437-963-103467	Sequence 103467,
30	37	77.1	969	16	US-10-437-963-165327	Sequence 165327,
31	36	75.0	73	12	US-10-424-599-249081	Sequence 249081,
32	36	75.0	155	12	US-10-424-599-190094	Sequence 190094,
33	36	75.0	284	12	US-10-425-114-55247	Sequence 55247, A
34	36	75.0	353	14	US-10-156-761-7849	Sequence 7849, Ap
35	35	72.9	10	9	US-09-848-585-19	Sequence 19, Appl
36	35	72.9	10	16	US-10-642-149-19	Sequence 19, Appl
37	35	72.9	43	12	US-10-424-599-209603	Sequence 209603,
38	35	72.9	46	14	US-10-029-386-29131	Sequence 29131, A
39	35	72.9	53	12	US-10-424-599-279107	Sequence 279107,
40	35	72.9	60	16	US-10-437-963-189501	Sequence 189501,
41	35	72.9	63	12	US-10-424-599-215072	Sequence 215072,
42	35	72.9	68	16	US-10-437-963-151038	Sequence 151038,
43	35	72.9	70	12	US-10-424-599-167020	Sequence 167020,
44	35	72.9	96	12	US-10-424-599-231295	Sequence 231295,
45	35	72.9	98	16	US-10-437-963-107817	Sequence 107817,

## ALIGNMENTS

RESULT 1  
US-10-615-959-41  
; Sequence 41, Application US/10615959  
; Publication No. US20040048321A1  
; GENERAL INFORMATION:  
; APPLICANT: Evre, David R.  
; TITLE OF INVENTION: SYNTHETIC PEPTIDES FOR COLLAGEN RESORPTION ASSAYS  
; FILE REFERENCE: WROS-1-18220  
; CURRENT FILING DATE: 2003-07-08  
; PRIOR FILING DATE: 2003-07-08  
; PRIOR FILING DATE: 2003-01-22  
; PRIOR FILING DATE: 1999-06-29  
; PRIOR FILING DATE: 1999-12-10  
; PRIOR FILING DATE: 1999-07-02  
; PRIOR FILING DATE: 1999-06-17  
; PRIOR FILING DATE: 1999-06-17  
; PRIOR FILING DATE: 1999-06-29  
; PRIOR FILING DATE: 1999-06-29  
; PRIOR FILING DATE: 1999-07-02  
; PRIOR FILING DATE: 1999-07-02  
; PRIOR FILING DATE: 1999-07-07  
; PRIOR FILING DATE: 1999-08-30  
; NUMBER OF SEQ ID NOS: 45  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 41  
; LENGTH: 9  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: MISC\_FEATURE  
; OTHER INFORMATION: synthetic peptide corresponding to carboxy-terminal  
; OTHER INFORMATION: telopeptide sequence of human type III collagen  
US-10-615-959-41

Query Match 100.0%; Score 48; DB 12; Length 9;  
Best Local Similarity 100.0%; Pred. No. 1.2e+06;

```

RESULT 3
US-10-366-125-11
; Sequence 11, Application US/10366125
; Publication No. US20030228259A1
; GENERAL INFORMATION:
; APPLICANT: Hellerstein, Marc
; TITLE OF INVENTION: MEASUREMENT OF BIOSYNTHESIS AND BREAKDOWN RATES OF
; TITLE OF INVENTION: BIOLOGICAL MOLECULES THAT ARE INACCESSIBLE OR NOT
; TITLE OF INVENTION: EASILY ACCESSIBLE TO DIRECT SAMPLING, NON-INVASIVELY,
; TITLE OF INVENTION: BY LABEL INCORPORATION INTO METABOLIC DERIVATIVES AND
; TITLE OF INVENTION: CATABOLITIC PRODUCTS
; FILE REFERENCE: 416272003500
; CURRENT APPLICATION NUMBER: US/10/366,125
; CURRENT FILING DATE: 2003-02-12
; PRIOR APPLICATION NUMBER: US 60/356,008
; PRIOR FILING DATE: 2002-02-12
; NUMBER OF SEQ ID NOS: 28
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 25
; TYPE: PRT

```

RESULT 5  
US-09-908-711-78  
; Sequence 78, Application US/09908711  
; Patent No. US20020045230A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: PA128  
; CURRENT APPLICATION NUMBER: US/09/908,711  
; CURRENT FILING DATE: 2001-07-20  
; PRIOR APPLICATION NUMBER: US01/01360  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: 09/764,857  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: US01/01344  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: 09/764,892  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: US01/01345  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: 09/764,898  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: US01/01329  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: 09/764,905  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: US01/01354  
; PRIOR FILING DATE: 2001-01-17



;; PRIOR APPLICATION NUMBER: 09/764,891  
;; PRIOR FILING DATE: 2001-01-17  
;; PRIOR APPLICATION NUMBER: US01/01339  
;; PRIOR FILING DATE: 2001-01-17  
;; PRIOR APPLICATION NUMBER: 09/764,869  
;; PRIOR FILING DATE: 2001-01-17  
;; PRIOR APPLICATION NUMBER: US01/01340  
;; PRIOR FILING DATE: 2001-01-17  
;; PRIOR APPLICATION NUMBER: 09/764,874  
;; PRIOR FILING DATE: 2001-01-17  
;; PRIOR APPLICATION NUMBER: US01/01334  
;; PRIOR FILING DATE: 2001-01-17  
;; PRIOR APPLICATION NUMBER: 09/764,898  
;; PRIOR FILING DATE: 2001-01-17  
;; PRIOR APPLICATION NUMBER: US01/01320  
;; PRIOR FILING DATE: 2001-01-17  
;; PRIOR APPLICATION NUMBER: 09/764,853  
;; PRIOR FILING DATE: 2001-01-17  
;; PRIOR APPLICATION NUMBER: US01/01349  
;; PRIOR FILING DATE: 2001-01-17  
;; PRIOR APPLICATION NUMBER: 09/764,902  
;; PRIOR FILING DATE: 2001-01-17  
;; PRIOR APPLICATION NUMBER: US01/01239  
;; PRIOR FILING DATE: 2001-01-17  
;; PRIOR APPLICATION NUMBER: 09/764,870  
;; PRIOR FILING DATE: 2001-01-17  
;; PRIOR APPLICATION NUMBER: US01/01348  
;; PRIOR FILING DATE: 2001-01-17  
;; PRIOR APPLICATION NUMBER: 09/764,882  
;; PRIOR FILING DATE: 2001-01-17  
;; PRIOR APPLICATION NUMBER: US01/01347  
;; PRIOR FILING DATE: 2001-01-17  
;; PRIOR APPLICATION NUMBER: 09/764,896  
;; PRIOR FILING DATE: 2001-01-17  
;; PRIOR APPLICATION NUMBER: US01/01307  
;; PRIOR FILING DATE: 2001-01-17  
;; PRIOR APPLICATION NUMBER: 09/764,864  
;; PRIOR FILING DATE: 2001-01-17  
;; PRIOR APPLICATION NUMBER: US01/01341  
;; PRIOR FILING DATE: 2001-01-17  
;; PRIOR APPLICATION NUMBER: 09/764,868  
;; PRIOR FILING DATE: 2001-01-17  
;; PRIOR APPLICATION NUMBER: US01/01312  
;; PRIOR FILING DATE: 2001-01-17  
;; PRIOR APPLICATION NUMBER: 60/179,065  
;; PRIOR FILING DATE: 2000-01-31  
;; PRIOR APPLICATION NUMBER: 60/180,628  
;; PRIOR FILING DATE: 2000-02-04  
;; PRIOR APPLICATION NUMBER: 60/209,467  
;; PRIOR FILING DATE: 2000-06-07  
;; NUMBER OF SEQ ID NOS: 167  
;; SOFTWARE: Patent In Ver. 2.0  
;; SEQ ID NO 78  
;; LENGTH: 309  
;; TYPE: PRT  
;; ORGANISM: Homo sapiens  
;; FEATURE:  
;; NAME/KEY: SITE  
;; LOCATION: (2)  
;; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
;; NAME/KEY: SITE  
;; LOCATION: (4)  
;; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
US-09-908-711-78

Query Match 100.0%; Score 48; DB 9; Length 309;  
Best Local Similarity 100.0%; Pred. No. 5.2;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IGGEKAGGF 9  
|||  
Db 51 IGGEKAGGF 59

## RESULT 6

US-09-918-715-226  
;; Sequence 226, Application US/09918715  
;; Publication No. US20030017157A1  
;; GENERAL INFORMATION:  
;; APPLICANT: Brad St. Croix  
;; APPLICANT: Bert Vogelstein  
;; APPLICANT: Kenneth Kinzler  
;; TITLE OF INVENTION: ENDOTHELIAL CELL EXPRESSION PATTERNS  
;; FILE REFERENCE: 1107.00134  
;; CURRENT APPLICATION NUMBER: US/09/918,715  
;; CURRENT FILING DATE: 2001-08-01  
;; PRIOR APPLICATION NUMBER: 60/222,599  
;; PRIOR FILING DATE: 2000-08-02  
;; PRIOR APPLICATION NUMBER: 60/224,360  
;; PRIOR FILING DATE: 2000-08-11  
;; PRIOR APPLICATION NUMBER: 60/282,850  
;; PRIOR FILING DATE: 2000-04-11  
;; NUMBER OF SEQ ID NOS: 358  
;; SOFTWARE: FastSeq for Windows Version 3.0  
;; SEQ ID NO 226  
;; LENGTH: 1466  
;; TYPE: PRT  
;; ORGANISM: Homo sapiens  
US-09-918-715-226

Query Match 100.0%; Score 48; DB 12; Length 1466;  
Best Local Similarity 100.0%; Pred. No. 23;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IGGEKAGGF 9  
|||  
Db 1208 IGGEKAGGF 1216

## RESULT 7

US-10-257-021-72  
;; Sequence 72, Application US/10257021  
;; Publication No. US20030211498A1  
;; GENERAL INFORMATION:  
;; APPLICANT: Morin, Patrice J.  
;; APPLICANT: Sherman-Baust, Cheryl A.  
;; APPLICANT: Pizer, Ellen S.  
;; APPLICANT: Hough, Colleen D.  
;; TITLE OF INVENTION: TUMOR MARKERS IN OVARIAN CANCER  
;; FILE REFERENCE: 14014.0369U2  
;; CURRENT APPLICATION NUMBER: US/10/257,021  
;; CURRENT FILING DATE: 2002-10-03  
;; PRIOR APPLICATION NUMBER: PCT/US01/10947  
;; PRIOR FILING DATE: 2001-04-03  
;; PRIOR APPLICATION NUMBER: 60/194,336  
;; PRIOR FILING DATE: 2000-04-03  
;; NUMBER OF SEQ ID NOS: 147  
;; SOFTWARE: FastSeq for Windows Version 4.0  
;; SEQ ID NO 72  
;; LENGTH: 1466  
;; TYPE: PRT  
;; ORGANISM: Homo sapiens  
US-10-257-021-72

Query Match 100.0%; Score 48; DB 12; Length 1466;  
Best Local Similarity 100.0%; Pred. No. 23;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IGGEKAGGF 9  
|||  
Db 1208 IGGEKAGGF 1216

```

; TITLE OF INVENTION: THERAPY OF COLON CANCER
; FILE REFERENCE: MPM01-029P2RNM
; CURRENT APPLICATION NUMBER: US/10/301,822
; CURRENT FILING DATE: 2002-11-21
; PRIOR APPLICATION NUMBER: US 60/339,971
; PRIOR FILING DATE: 2001-12-10
; PRIOR APPLICATION NUMBER: US 60/361,978
; PRIOR FILING DATE: 2002-03-05
; PRIOR APPLICATION NUMBER: US 60/381,988
; PRIOR FILING DATE: 2002-05-20
; NUMBER OF SEQ ID NOS: 228
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 33
; LENGTH: 1466
; TYPE: PRT
; ORGANISM: Homo Sapiens
US-10-301-822-33

Query Match 100.0%; Score 48; DB 14; Length 1466;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 IGGEKAGGF 9
Db 1208 IGGEKAGGF 1216

RESULT 10
US-10-357-851-3
; Sequence 3, Application US/10357851
; Publication No. US20040151731A1
; GENERAL INFORMATION:
; APPLICANT: Jicha, Douglas L.
; TITLE OF INVENTION: Method and Compositions Involving
; T-Lymphocyte Reactivity with Collagen in Aortic Tissue of
; Abdominal Aortic Aneurysm Patients
; FILE REFERENCE: 13376US
; CURRENT APPLICATION NUMBER: US/10/357,851
; CURRENT FILING DATE: 2003-02-04
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 1466
; TYPE: PRT
; ORGANISM: Homo Sapiens
US-10-357-851-3

Query Match 100.0%; Score 48; DB 16; Length 1466;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 IGGEKAGGF 9
Db 1208 IGGEKAGGF 1216

RESULT 11
US-10-358-024-3
; Sequence 3, Application US/10358024
; Publication No. US20040151732A1
; GENERAL INFORMATION:
; APPLICANT: Pelaez, Stephen
; APPLICANT: Jicha, Douglas L.
; TITLE OF INVENTION: Methods and Compositions Involving Blood
; T-Lymphocytes Reactivity with Collagen
; FILE REFERENCE: 13436US
; CURRENT APPLICATION NUMBER: US/10/358,024
; CURRENT FILING DATE: 2003-02-04
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 1466
; TYPE: PRT

```

```

; TITLE OF INVENTION: COMPOSITIONS, KITS, AND METHODS FOR IDENTIFICATION, ASSESSMENT,
; PREVENTION, AND THERAPY OF BREAST CANCER
; FILE REFERENCE: MRI-038
; CURRENT APPLICATION NUMBER: US/10/177,293
; CURRENT FILING DATE: 2002-06-21
; PRIOR APPLICATION NUMBER: US 60/299,887
; PRIOR FILING DATE: 2001-06-21
; PRIOR APPLICATION NUMBER: US 60/301,572
; PRIOR FILING DATE: 2001-06-27
; PRIOR APPLICATION NUMBER: US 60/306,501
; PRIOR FILING DATE: 2001-07-18
; PRIOR APPLICATION NUMBER: US 60/325,002
; PRIOR FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: US 60/362,585
; PRIOR FILING DATE: 2002-03-05
; PRIOR APPLICATION NUMBER: US 60/xxx,xxx
; PRIOR FILING DATE: 2002-05-14
; NUMBER OF SEQ ID NOS: 506
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 68
; LENGTH: 1466
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-177-293-68

Query Match 100.0%; Score 48; DB 14; Length 1466;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 IGGEKAGGF 9
Db 1208 IGGEKAGGF 1216

RESULT 9
US-10-301-822-33
; Sequence 33, Application US/10301822
; Publication No. US20030148410A1
; GENERAL INFORMATION:
; APPLICANT: Millennium Pharmaceuticals, Inc.
; APPLICANT: Berger, Allison
; APPLICANT: Guillemette, Tracy L.
; APPLICANT: Kamatkar, Shubhangi
; APPLICANT: Schlegel, Robert
; APPLICANT: Monahan, John E.
; APPLICANT: Thibodeau, Stephen N.
; APPLICANT: Burgart, Lawrence J.
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND
; METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND

```

```
; ORGANISM: Homo Sapiens
US-10-358-024-3

Query Match      100.0%; Score 48; DB 16; Length 1466;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 IGGEKAGGF 9
Db      1208 IGGEKAGGF 1216

RESULT 12
US-10-734-564-103
; Sequence 103, Application US/10734564
; Publication No. US20040157278A1
; GENERAL INFORMATION:
; APPLICANT: Christopher C Burgess et al
; TITLE OF INVENTION: Detection Methods Using TIMP1
; FILE REFERENCE: 1657/2012
; CURRENT APPLICATION NUMBER: US/10/734,564
; CURRENT FILING DATE: 2003-12-12
; NUMBER OF SEQ ID NOS: 138
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 103
; LENGTH: 1466
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-734-564-103

Query Match      100.0%; Score 48; DB 16; Length 1466;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 IGGEKAGGF 9
Db      1208 IGGEKAGGF 1216

RESULT 13
US-10-402-089-12
; Sequence 12, Application US/10402089
; Publication No. US20040005663A1
; GENERAL INFORMATION:
; APPLICANT: Bell, Marcum P.
; APPLICANT: Neff, Thomas B.
; APPLICANT: Polarek, James W.
; TITLE OF INVENTION: PORCINE COLLAGENS AND GELATINS
; FILE REFERENCE: FP0402.3 CON
; CURRENT APPLICATION NUMBER: US/10/402,089
; CURRENT FILING DATE: 2003-03-26
; PRIOR APPLICATION NUMBER: US 09/709,700
; PRIOR FILING DATE: 2000-11-10
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 12
; LENGTH: 1466
; TYPE: PRT
; ORGANISM: Sus scrofa
US-10-402-089-12

Query Match      97.9%; Score 47; DB 15; Length 1466;
Best Local Similarity 88.9%; Pred. No. 34;
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      1 IGGEKAGGF 9
Db      1208 VGGEKAGGF 1216

RESULT 14
US-10-402-072A-12
; Sequence 12, Application US/10402072A
; Publication No. US20040018592A1
; GENERAL INFORMATION:
; APPLICANT: Bell, Marcum P.
; APPLICANT: Neff, Thomas B.
; APPLICANT: Polarek, James W.
; APPLICANT: Seeley, Todd W.
; TITLE OF INVENTION: BOVINE COLLAGENS AND GELATINS
; FILE REFERENCE: FP0402.2 CON
; CURRENT APPLICATION NUMBER: US/10/402,072A
; CURRENT FILING DATE: 2003-03-26
; PRIOR APPLICATION NUMBER: US 09/709,700
; PRIOR FILING DATE: 2000-11-10
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 12
; LENGTH: 1466
; TYPE: PRT
; ORGANISM: Sus scrofa
US-10-402-072A-12

Query Match      97.9%; Score 47; DB 15; Length 1466;
Best Local Similarity 88.9%; Pred. No. 34;
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      1 IGGEKAGGF 9
Db      1208 VGGEKAGGF 1216

RESULT 15
US-10-615-959-40
; Sequence 40, Application US/10615959
; Publication No. US20040048321A1
; GENERAL INFORMATION:
; APPLICANT: Eyre, David R.
; TITLE OF INVENTION: SYNTHETIC PEPTIDES FOR COLLAGEN RESORPTION ASSAYS
; FILE REFERENCE: WROS-1-18220
; CURRENT APPLICATION NUMBER: US/10/615,959
; CURRENT FILING DATE: 2003-07-08
; PRIOR APPLICATION NUMBER: US/10/009,999A
; PRIOR FILING DATE: 2003-01-22
; PRIOR APPLICATION NUMBER: PCT/US99/29357
; PRIOR FILING DATE: 1999-12-10
; PRIOR APPLICATION NUMBER: US 09/335,098
; PRIOR FILING DATE: 1999-06-17
; PRIOR APPLICATION NUMBER: US 60/141,574
; PRIOR FILING DATE: 1999-06-29
; PRIOR APPLICATION NUMBER: US 60/142,274
; PRIOR FILING DATE: 1999-07-02
; PRIOR APPLICATION NUMBER: US 60/142,675
; PRIOR FILING DATE: 1999-07-07
; PRIOR APPLICATION NUMBER: US 09/385,740
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 40
; LENGTH: 8
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: MISC FEATURE
; OTHER INFORMATION: synthetic peptide corresponding to carboxy-terminal
; OTHER INFORMATION: telopeptide sequence of human type III collagen
US-10-615-959-40

Query Match      87.5%; Score 42; DB 12; Length 8;
Best Local Similarity 100.0%; Pred. No. 1.2e+06;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 IGGEKAGG 8
Db      1 IGGEKAGG 8
```

Mon Sep 20 11:05:44 2004

Search completed: September 18, 2004, 04:56:13  
Job time : 43.7059 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 18, 2004, 04:02:05 ; Search time 10.9412 Seconds  
(without alignments)  
79.125 Million cell updates/sec

Title: US-10-615-959-41

Perfect score: 48

Sequence: 1 IGGEXAGGF 9

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR\_78:\*

1: pir1:\*

2: pir2:\*

3: pir3:\*

4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	48	100.0	1466	1 CGHU7L	collagen alpha 1(I
2	44	91.7	636	2 S41067	collagen alpha 1(I
3	44	91.7	1464	2 S59856	collagen alpha 1(I
4	38	79.2	383	2 G72777	probable S2P metal
5	37	77.1	476	2 AG1031	hypothetical prote
6	36	75.0	357	2 AH2756	hypothetical prote
7	36	75.0	379	2 G97537	glycine cleavage s
8	36	75.0	404	2 S56535	hypothetical 43.7K
9	36	75.0	404	2 E91287	hypothetical prote
10	36	75.0	404	2 H86128	hypothetical prote
11	36	75.0	629	2 A29666	keratin, 65K type
12	36	75.0	629	2 S42629	keratin K3 - rabbi
13	35	72.9	161	2 H4998	heat shock protein
14	35	72.9	270	2 H69074	formylmethanofuran
15	35	72.9	270	2 S57457	formylmethanofuran
16	35	72.9	324	2 E84505	N-acetylmuramoyl-L
17	35	72.9	375	2 AF1720	probable NADPH deh
18	35	72.9	382	2 S55482	hypothetical prote
19	35	72.9	440	2 T24323	hypothetical prote
20	35	72.9	457	2 B87269	hypothetical prote
21	35	72.9	468	2 A45268	interleukin-9 rece
22	35	72.9	475	2 AE2301	phosphoglucomutase
23	35	72.9	490	2 S75539	hypothetical prote
24	35	72.9	606	2 G72282	hypothetical prote
25	35	72.9	1009	2 S28081	pol polyprotein -
26	35	72.9	1028	2 B24785	hypothetical prote
27	35	72.9	1440	2 C94639	hypothetical prote
28	34	70.8	165	2 T41103	60s ribosomal prot
29	34	70.8	168	2 H69149	conserved hypothet

30	34	70.8	240	2 B36724	hypothetical prote
31	34	70.8	266	2 T31264	cis-1,2-dihydro-1,
32	34	70.8	276	2 T33925	hypothetical prote
33	34	70.8	324	2 T14636	hypothetical prote
34	34	70.8	371	2 G96937	hypothetical prote
35	34	70.8	457	2 C84378	cell wall-associat
36	34	70.8	470	2 F82302	phosphoglucomutase
37	34	70.8	499	2 C70345	probable phosphogl
38	34	70.8	561	2 S62788	phosphoglucomutase
39	34	70.8	561	2 S71597	carboxylesterase (
40	34	70.8	565	2 S10367	carboxylesterase (
41	34	70.8	608	2 T32923	hypothetical prote
42	34	70.8	644	2 T02366	hypothetical prote
43	34	70.8	718	2 H86309	hypothetical prote
44	34	70.8	948	2 C82077	F28G4.21 protein -
45	34	70.8	1137	1 WMBEB1	glutamate-ammonia- ribonucleoside-dip

## ALIGNMENTS

### RESULT 1

CGHU7L

collagen alpha 1(III) chain precursor - human

N;Alternate names: procollagen alpha 1(III) chain

C;Species: Homo sapiens (man)

C;Date: 24-Apr-1984 #sequence revision 01-Sep-1995 #text change 21-Jul-2000

C;Accession: S05272; S04642; PE0011; S01726; S04887; A90399; A94562; I51868; S59511; A904

R;Prockop, D.J.

submitted to the EMBL Data Library, February 1989

A;Reference number: S05272

A;Accession: S05272

A;Status: preliminary

A;Molecule type: mRNA

A;Residues: 1-1240,'V',1242-1466 <PRC>

A;Cross-references: EMBL:X14420; NID:G30057; PIDN:CAA32583.1; PID:G30058

R;Ala-Kokko, L.; Kontusaari, S.; Baldwin, C.T.; Kuivaniemi, H.; Prockop, D.J.

Biochem. J. 260, 509-516, 1989

A;Title: Structure of cDNA clones coding for the entire prepro-alpha(III) chain of human

erences.

A;Reference number: S04642; MUID:89350838; PMID:2764886

A;Accession: S04642

A;Molecule type: mRNA

A;Residues: 1-1196 <ALA>

A;Cross-references: EMBL:X14420; NID:G30057; PIDN:CAA32583.1; PID:G30058

A;Note: the complete sequence is not shown

R;Benson-Chanda, V.; Su, M.W.; Weil, D.; Chu, M.L.; Ramirez, F.

Gene 78, 255-265, 1989

A;Title: Cloning and analysis of the 5' portion of the human type-III procollagen gene ((

A;Reference number: PE0011; MUID:89378752; PMID:2777083

A;Accession: PE0011

A;Molecule type: DNA

A;Residues: 1-176 <BEN>

A;Cross-references: NID:G180813; PIDN:AAA52040.1; PID:G180814

R;Toman, P.D.; Ricca, G.A.; de Crombrughe, B.

Nucleic Acids Res. 16, 7201, 1988

A;Title: Nucleotide sequence of a cDNA coding for the amino-terminal region of human prei

A;Reference number: S01726; MUID:88303360; PMID:3405773

A;Accession: S01726

A;Molecule type: mRNA

A;Residues: 1-170 <TOM>

A;Cross-references: EMBL:X07240; NID:G30050; PIDN:CAA30229.1; PID:G30061

A;Note: the authors translated the codon CAG for residue 154 as His

R;Janeczko, R.A.; Ramirez, F.

Nucleic Acids Res. 17, 6742, 1989

A;Title: Nucleotide and amino acid sequences of the entire human alpha-1 (III) collagen.

A;Reference number: S04887; MUID:89386015; PMID:2780304

A;Accession: S04887

A;Molecule type: mRNA

A;Residues: 149-163,'G',164-240,'D',242-471,'D',473-487,'L',489,'S',491-613,'Y',615-634,'

A;Cross-references: EMBL:X15332; NID:G29545; PIDN:CAA33387.1; PID:G930045

A;Note: the authors' translation of residues 905-932 is inconsistent with the nucleotide

R;Seyer, J.M.; Kang, A.H.

Mon Sep 20 11:05:45 2004

Biochemistry 16, 1158-1164, 1977  
A>Title: Covalent structure of collagen: amino acid sequence of cyanogen bromide peptide  
A:Reference number: A90399; MUID:77134724; PMID:557335  
A:Accession: A90399  
A:Molecule type: protein  
A:Residues: 'V', 169-225, 229-232, 'P', 234-292, 'D', 294-398 <SEY1>  
A:Experimental source: liver  
A>Note: sequence corrected by A94562; attachment of 2-O-alpha-D-glucosyl-O-beta-D-galactose  
R:Seyer, J.M.  
Submitted to the Atlas December 1977  
A:Reference number: A94562  
A:Accession: A94562  
A:Molecule type: protein  
A:Residues: 'V', 169-225, 229-277, 'A', 279-292, 'D', 294, 'S', 296-398 <SEY2>  
A:Experimental source: liver  
A>Note: author submitted corrections to A90399  
R:Milewicz, D.M.; Witz, A.M.; Smith, A.C.; Manchester, D.K.; Waldstein, G.; Byers, P.H.  
Am. J. Hum. Genet. 53, 62-70, 1993  
A>Title: Parental somatic and germ-line mosaicism for a multiexon deletion with unusual  
A:Reference number: I51868; MUID:93304430; PMID:8317500  
A:Accession: I51868  
A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 186-194 <MIL>  
A:Cross-references: GB:S62925; NID:g386425; PIDN:AAAD1937.1; PID:g4261637  
R:Chiodo, A.A.; Sillence, D.O.; Cole, W.G.; Bateman, J.F.  
Biochem. J. 311, 939-943, 1995  
A>Title: Abnormal type III collagen produced by an exon-17-skipping mutation of the COL3A1  
A:Reference number: S59511; MUID:96067614; PMID:7487954  
A:Accession: S59511  
A:Molecule type: mRNA  
A:Residues: 302-423 <CHI>  
A:Cross-references: GB:S79877; NID:g1195576; PIDN:AAB35615.1; PID:g1195577  
R:Seyer, J.M.; Kang, A.H.  
Biochemistry 17, 3404-3411, 1978  
A>Title: Covalent structure of collagen: amino acid sequence of five consecutive CNBr peptides  
A:Reference number: A90414; MUID:79000343; PMID:687591  
A:Accession: A90414  
A:Molecule type: protein  
A:Residues: 399-675, 'N', 677-727 <SEY3>  
A:Experimental source: liver  
R:Lee, B.; Vitale, E.; Superti-Furga, A.; Steinmann, B.; Ramirez, F.  
J. Biol. Chem. 266, 5256-5259, 1991  
A>Title: G to T transversion at position +5 of a splice donor site causes skipping of the  
A:Reference number: I55349; MUID:91161621; PMID:1672129  
A:Accession: I55349  
A>Status: translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 537-605 <LEE>  
A:Cross-references: GB:M59312; NID:g180815; PIDN:AAA52041.1; PID:g180816  
R:Seyer, J.M.; Mainardi, C.; Kang, A.H.  
Biochemistry 19, 1583-1589, 1980  
A>Title: Covalent structure of collagen: amino acid sequence of alpha1 (III)-CB5 from the  
A:Reference number: A90438; MUID:80198282; PMID:6246925  
A:Accession: A90438  
A:Molecule type: protein  
A:Residues: 728-895, 'A', 897-964 <SEY4>  
A:Experimental source: liver  
R:Coile, W.G.; Chiodo, A.A.; Laman, S.R.; Janeczko, R.; Ramirez, F.; Dahl, H.H.M.; Char  
J. Biol. Chem. 265, 17070-17077, 1990  
A>Title: A base substitution at a splice site in the COL3A1 gene causes exon skipping and  
A:Reference number: A38303; MUID:91009133; PMID:2145268  
A:Accession: A38303  
A:Molecule type: mRNA  
A:Residues: 861-1015 <COL>  
A:Cross-references: GB:J05617; GB:M55603; GB:M59227; NID:g180878; PIDN:AAB59383.1; PID:g  
R:Mankoo, B.S.; Dalglish, R.  
Nucleic Acids Res. 16, 2337, 1988  
A>Title: Human pro alpha1(III) collagen: cDNA sequence for the 3' end.  
A:Reference number: S02119; MUID:88189827; PMID:3357782  
A:Accession: S02119

A>Status: translation not shown  
A:Molecule type: mRNA  
A:Residues: 950-1018, 'Y', 1020-1193, 'S', 1185-1466 <MAN>  
A:Cross-references: EMBL:X06700; NID:g30053; PIDN:CAA29886.1; PID:g30054  
R:Seyer, J.M.; Kang, A.H.  
Biochemistry 20, 2621-2627, 1981  
A>Title: Covalent structure of collagen: amino acid sequence of alpha1 (III)-CB9 from the  
A:Reference number: A90446; MUID:81208139; PMID:7016180  
A:Accession: A90446  
A:Molecule type: protein  
A:Residues: 965-979, 'A', 981-984, 'PS', 987, 'QN', 990-1096, 'P', 1098-1152, 'AT', 1155, 'S', 1157-1  
A:Experimental source: liver  
R:Loidl, H.R.; Brinker, J.M.; May, M.; Pihlajaniemi, T.; Morrow, S.; Rosenbloom, J.; Myer  
Nucleic Acids Res. 12, 9383-9394, 1984  
A>Title: Molecular cloning and carboxyl-propeptide analysis of human type III procollagen  
A:Reference number: A93551; MUID:85087944; PMID:6096827  
A:Accession: A93551  
A:Molecule type: mRNA  
A:Residues: 1065-1155, 'P', 1157-1466 <LOI>  
A:Cross-references: EMBL:X01655; EMBL:X01742; NID:g29584; PIDN:CAA25821.1  
R:Miskulin, M.; Dalglish, R.; Kluge-Becker, B.; Rennard, S.I.; Tolstoshev, P.; Brant  
Biochemistry 25, 1408-1413, 1986  
A>Title: Human type III collagen gene expression is coordinately modulated with the type  
A:Reference number: I52393; MUID:86187804; PMID:375462  
A:Accession: I52393  
A>Status: translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1161-1200 <MIS>  
A:Cross-references: GB:M13146; NID:g180415; PIDN:AAA52003.1; PID:g180416  
R:Emanuel, B.S.; Cannizzaro, L.A.; Seyer, J.M.; Myers, J.C.  
Proc. Natl. Acad. Sci. U.S.A. 82, 3385-3389, 1985  
A>Title: Human alpha 1(III) and alpha 2(V) procollagen genes are located on the long arm  
A:Reference number: I59025; MUID:85216505; PMID:3858826  
A:Accession: I59025  
A>Status: translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1165-1196 <EMA>  
A:Cross-references: GB:M11134; NID:g180417; PIDN:AAA52004.1; PID:g180418  
R:Chu, M.L.; Weil, D.; de Wet, W.; Bernard, M.; Sippola, M.; Ramirez, F.  
J. Biol. Chem. 260, 4357-4363, 1985  
A>Title: Isolation of cDNA and genomic clones encoding human pro-alpha1(III) collagen.  
A:Reference number: A92516; MUID:85157600; PMID:2579994  
A:Accession: A92516  
A:Molecule type: DNA  
A:Residues: 1176-1240, 'V', 1242-1356, 'P', 1358-1466 <CHU>  
A:Cross-references: GB:M10615; GB:M10793; GB:M10794; GB:M10795; GB:M10796; GB:M10797; GB:M10798  
A:Experimental source: liver  
A>Note: the authors translated the codon TTC for residue 1057 as Tyr; the codons given f  
ation  
C:Comment: Prolines and lysines at the third position of the tripeptide repeating unit (3-hydroxylated. About 15% of the lysines are 5-hydroxylated and some are subsequently O  
C:Genetics:  
A:Gene: GDB:COL3A1  
A:Cross-references: GDB:118729; OMIM:120180  
A:Map position: 2q31-2q31  
A:Introns: 27/1, 94/3, 111/3, 149/3, 176/3, 554/3, 587/3, 1175/3, 1275/3, 1337/3, 1418/3  
A>Note: the list of introns is incomplete; defects in this gene can result in Ehlers-Dan  
C:Complex: type III collagen is a homotrimer of monomers initially linked by disulfide b  
er of their length, is formed with desmosine cross-links made from lysine and allysine r  
C:Function:  
A:Description: structural component of extracellular fibrous polymer that maintains inte  
C:Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;  
C:Keywords: coiled coil; Ehlers-Danlos syndrome; extracellular matrix; glycoprotein; hyd  
F1-23/Domain: signal sequence #status predicted <SIG>  
F124-153/Domain: amino-terminal propeptide #status predicted <PRO>  
F31-91/Domain: von Willebrand factor type C repeat homology <VWC>  
F154-1221/Product: collagen alpha 1(III) chain #status predicted <MAT>  
F154-167/Region: amino-terminal nonhelical telopeptide  
F168-1196/Region: helical  
F1091-1093/Region: cell attachment (R-G-D) motif  
F1197-1221/Region: carboxyl-terminal nonhelical telopeptide  
F1222-1466/Domain: carboxyl-terminal propeptide #status predicted <CPR>  
F1238-1466/Domain: fibrillar collagen carboxyl-terminal homology <PCC>

A:Accession: S59856  
A:Molecule type: DNA  
A:Residues: 1-1464 <TOW>  
A:Cross-references: EMBL:X52046  
R:Toman, D.  
submitted to the EMBL Data Library, November 1994  
A:Reference number: S62120  
A:Accession: S62120  
A:Molecule type: DNA  
A:Residues: 1-866, 'G', 868-1464 <TOA>  
R:Metaseraanta, M.; Toman, D.; de Crombrughe, B.; Vuorio, E.  
Biochim. Biophys. Acta 1089, 241-243, 1991  
A:Title: Specific hybridization probes for mouse type I, II, III and IX collagen mRNAs.  
A:Reference number: S16176; MUID:91274355; PMID:2054384  
A:Accession: S16373  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1442-1464 <MET>  
A:Cross-references: EMBL:X57983; NID:G50476; PIDN:CAA41048.1; PID:G50477  
C:Genetics:  
A:Introns: 29/1; 95/3; 112/3; 150/3; 175/3; 193/3; 211/3; 229/3; 247/3; 265/3; 283/3; 293/3;  
58/3; 673/3; 706/3; 742/3; 760/3; 778/3; 786/3; 814/3; 850/3; 868/3; 886/3; 940/3; 976/3;  
C:Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;  
C:Keywords: coiled coil; extracellular matrix  
F:1-24/Domain: signal sequence #status predicted <SIG>  
F:25-154/Domain: propeptide #status predicted <PRO>  
F:32-92/Domain: von Willebrand factor type C repeat homology <VWC>  
F:155-1464/Product: collagen alpha 1(III) chain #status predicted <MAT>  
F:1236-1464/Domain: fibrillar collagen carboxyl-terminal homology <PCC>

Query Match 91.7%; Score 44; DB 2; Length 1464;  
Best Local Similarity 77.8%; Pred. No. 8.3;  
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 IGGEKAGGF 9  
Db 1206 VGGEKSGGF 1214  
:||||:|

RESULT 4  
G72777  
Probable S2P metalloproteinase APE0209 - Aeropyrum pernix (strain K1)  
C:Species: Aeropyrum pernix  
C:Date: 20-Aug-1999 #sequence\_revision 20-Aug-1999 #text\_change 20-Aug-1999  
C:Accession: G72777  
R:Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Takahara, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; K.  
DNA Res. 6, 83-103, 1999  
A:Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyrum  
A:Reference number: A72450; MUID:99310339; PMID:10382966  
A:Accession: G72777  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-383 <KAW>  
A:Cross-references: DDBJ:AP000058; NID:G5103388; PIDN:BAA79121.1; PID:d1042897; PID:G5103388  
A:Experimental source: strain K1  
C:Genetics:  
A:Gene: APE0209

Query Match 79.2%; Score 38; DB 2; Length 383;  
Best Local Similarity 87.5%; Pred. No. 28;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 GGEKAGGF 9  
Db 99 GGEAAGGF 106  
|||||

RESULT 5  
AG1031  
hypothetical protein STY4577 [imported] - Salmonella enterica subsp. enterica serovar Typhimurium  
C:Species: Salmonella enterica subsp. enterica serovar Typhimurium

A;Note: this species has also been called *Salmonella typhi*  
C;Date: 09-Nov-2001 #sequence\_revision 09-Nov-2001 #text\_change 18-Nov-2002  
C;Accession: AG1031  
R;Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher, T.; Connor, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar, S.; Moule, S.; O'Gaora, P.  
Nature 413, 848-852, 2001  
A;Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; A;Title: Complete genome sequence of a multiple drug resistant *Salmonella enterica* serovar  
A;Reference number: AB0502; MUID:21534947; PMID:11677608  
A;Accession: AG1031  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-476 <PAR>  
A;Cross-references: GB:AL513382; PIDN:CAD09352.1; PID:gl6505352; GSPDB:GN00176  
C;Genetics:  
A;Gene: STY4577

Query Match 77.1%; Score 37; DB 2; Length 476;  
Best Local Similarity 75.0%; Pred. No. 52;  
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 IGGEKAGG 8  
Db 211 VGGKAGG 218  
:||:||||  
:||:||||

RESULT 6  
AH2756  
hypothetical protein gcvT [imported] - Agrobacterium tumefaciens (strain C58, Dupont)  
C;Species: Agrobacterium tumefaciens  
C;Date: 11-Jan-2002 #sequence\_revision 11-Jan-2002 #text\_change 18-Nov-2002  
C;Accession: AH2756  
R;Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I.; Erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; McClellan, X.P.; Romero, P.; Zhang, S.  
Science 294, 2317-2323, 2001  
A;Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm, S.; E.W.  
A;Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.  
A;Reference number: AB2577; MUID:21608550; PMID:11743193  
A;Accession: AH2756  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-357 <KUR>  
A;Cross-references: GB:AB008688; PIDN:AAL42470.1; PID:gl7739886; GSPDB:GN00186  
A;Experimental source: strain C58 (Dupont)  
C;Genetics:  
A;Gene: gcvT  
A;Map position: circular chromosome  
C;Superfamily: aminomethyltransferase

Query Match 75.0%; Score 36; DB 2; Length 357;  
Best Local Similarity 75.0%; Pred. No. 60;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 GGEKAGGF 9  
Db 248 GGEREGGF 255  
||||:||||  
||||:||||

RESULT 7  
G97537  
glycine cleavage system protein T2 (PA2442) [imported] - Agrobacterium tumefaciens (strain C58)  
C;Species: Agrobacterium tumefaciens  
C;Date: 30-Sep-2001 #sequence\_revision 30-Sep-2001 #text\_change 18-Nov-2002  
C;Accession: G97537  
R;Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman, A.; Liu, F.; Wollam, C.; Allinger, M.; Doughy, D.; Scott, C.; Lappas, C.; Markelz, B.; Science 294, 2323-2328, 2001  
A;Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tumefaciens C58  
A;Reference number: A97359; MUID:21608551; PMID:11743194  
A;Accession: G97537

A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-379 <KUR>  
A;Cross-references: GB:AE007869; PIDN:AAK87256.1; PID:gl5156544; GSPDB:GN00169  
C;Genetics:  
A;Gene: AGR C 2701  
A;Map position: circular chromosome  
C;Superfamily: aminomethyltransferase

Query Match 75.0%; Score 36; DB 2; Length 379;  
Best Local Similarity 75.0%; Pred. No. 64;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 GGEKAGGF 9  
Db 270 GGEREGGF 277  
||||:||||  
||||:||||

RESULT 8  
S56535  
hypothetical 43.7K protein (fecI-fimb intergenic region) - Escherichia coli (strain K-12)  
N;Alternate names: hypothetical protein f404  
C;Species: Escherichia coli  
C;Date: 28-Oct-1995 #sequence\_revision 03-Nov-1995 #text\_change 01-Mar-2002  
C;Accession: S56535; H65244  
R;Burland, V.; Plunkett III, G.; Sofia, H.J.; Daniels, D.L.; Blattner, F.R.  
Nucleic Acids Res. 23, 2105-2119, 1995  
A;Title: Analysis of the Escherichia coli genome VI: DNA sequence of the region from 92.6.8.1 to 92.6.8.10.1  
A;Reference number: S56314; MUID:95334362; PMID:7610040  
A;Accession: S56535  
A;Status: preliminary; nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-404 <BUR>  
A;Cross-references: EMBL:U14003; NID:gl263172; PIDN:AAA97206.1; PID:g537151  
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, August 1994  
R;Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.F.; Burland, V.; Riley, M.; A.; Rose, D.J.; Mau, B.; Shao, Y.  
Science 277, 1453-1462, 1997  
A;Title: The complete genome sequence of Escherichia coli K-12.  
A;Reference number: S54720; MUID:97426617; PMID:9278503  
A;Accession: H65244  
A;Status: preliminary; nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-404 <BLAT>  
A;Cross-references: GB:AE000501; GB:U00096; NID:g2367372; PIDN:AAC77266.1; PID:gl790764;  
A;Experimental source: strain K-12, substrain MG1655  
C;Genetics:  
A;Gene: Yjht

Query Match 75.0%; Score 36; DB 2; Length 404;  
Best Local Similarity 87.5%; Pred. No. 68;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 IGGEKAGG 8  
Db 377 IGGETAGG 384  
||||:||||  
||||:||||

RESULT 9  
E91287  
hypothetical protein Ecs5269 [imported] - Escherichia coli (strain O157:H7, substrain R1)  
C;Species: Escherichia coli  
C;Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 18-Jul-2001  
C;Accession: E91287  
R;Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.; Gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.  
DNA Res. 8, 11-22, 2001  
A;Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genomic  
A;Reference number: A99629; MUID:21156231; PMID:11258796  
A;Accession: E91287  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-404 <HAY>



A:Cross-references: GB:BA000007; PIDN:BA838692.1; PID:gl3364747; GSPDB:GNO0154  
A:Experimental source: strain O157:H7, substrain RMD 0509952  
C:Genetics:  
A:Gene: ECs5269

Query Match 75.0%; Score 36; DB 2; Length 404;  
Best Local Similarity 87.5%; Pred. No. 68;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 IGGEKAGG 8  
||| |||  
Db 377 IGGETAGG 384

RESULT 10  
H6128  
hypothetical protein yjht [imported] - Escherichia coli (strain O157:H7, substrain EDL933)  
C:Species: Escherichia coli  
C>Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 14-Sep-2001  
C:Accession: H86128  
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew  
Miller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,  
Nature 409, 529-533, 2001  
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.  
A:Reference number: A85480; MUID:21074935; PMID:11206551  
A:Accession: H86128  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-404 <STO>  
A:Cross-references: GB:AE005174; NID:gl2519317; PIDN:AA059492.1; GSPDB:GNO0145; UWGP:259  
A:Experimental source: strain O157:H7, substrain EDL933  
C:Genetics:  
A:Gene: yjht

Query Match 75.0%; Score 36; DB 2; Length 404;  
Best Local Similarity 87.5%; Pred. No. 68;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 IGGEKAGG 8  
||| |||  
Db 377 IGGETAGG 384

RESULT 11  
A29666  
keratin, 65K type II cytoskeletal - human  
C:Species: Homo sapiens (man)  
C>Date: 31-Dec-1988 #sequence\_revision 30-Sep-1991 #text\_change 20-Apr-2001  
C:Accession: A29666  
R:Klinge, E.M.; Sylvestre, Y.R.; Freedberg, I.M.; Blumenberg, M.  
J. Mol. Evol. 24, 319-329, 1987  
A:Title: Evolution of keratin genes: different protein domains evolve by different pathw  
A:Reference number: A29666; MUID:87254239; PMID:2439698  
A:Accession: A29666  
A:Molecule type: DNA  
A:Residues: 1-629 <KLI>  
A:Cross-references: GB:X05418; NID:g34040; PIDN:CAA28991.1; PID:g34041; GB:X05419; NID:g  
D:gl335193; PIDN:CAA28995.1; PID:gl335194; GB:X05421; NID:g34047; PIDN:CAA28996.1; PID:g  
A:Note: the authors translated the codon AAC for residue 63 as Asp and ACA for residue 2  
A:Note: the complete translation is not annotated in GenBank entries HSKER65A, HSKER65B,  
s mistranslated as a Met initiator codon  
C:Genetics:  
A:Introns: 215/3; 289/2; 309/3; 341/3; 396/3; 438/3; 512/2; 525/1  
A:Superfamily: cytoskeletal keratin  
C:Keywords: coiled coil

Query Match 75.0%; Score 36; DB 2; Length 629;  
Best Local Similarity 75.0%; Pred. No. 14+02;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 GGEKAGG 9  
||| |||  
Db 73 GGSRAGG 80

RESULT 12  
S42629

keratin K3 - rabbit  
C:Species: Oryctolagus cuniculus (domestic rabbit)  
C>Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 13-Aug-1999  
C:Accession: S42629  
R:Wu, R.L.; Galvin, S.; Wu, S.K.; Xu, C.; Blumenberg, M.; Sun, T.T.  
J. Cell Sci. 105, 303-316, 1993  
A:Title: A 300 bp 5'-upstream sequence of a differentiation-dependent rabbit K3 keratin  
A:Reference number: S42629; MUID:94013038; PMID:7691837  
A:Accession: S42629  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-629 <WUR>  
A:Cross-references: EMBL:X74371; NID:g433957; PIDN:CAA52409.1; PID:g433958  
C:Genetics:  
A:Introns: 200/3; 274/2; 294/3; 326/3; 381/3; 423/3; 497/2; 509/1  
C:Superfamily: cytoskeletal keratin

Query Match 75.0%; Score 36; DB 2; Length 629;  
Best Local Similarity 75.0%; Pred. No. 14+02;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 GGEKAGG 9  
||| |||  
Db 79 GGSRAGG 86

RESULT 13

S14998  
heat shock protein 18 (clone c9) - maize

C:Species: Zea mays (maize)  
C>Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 13-Aug-1999  
C:Accession: S14998  
R:Coping, I.S.; Frappier, J.R.H.; Walden, D.B.; Atkinson, B.G.  
Plant Mol. Biol. 16, 699-711, 1991  
A:Title: Sequence, identification and characterization of cDNAs encoding two different me  
A:Reference number: S14997; MUID:91329703; PMID:1714322  
A:Accession: S14998  
A:Molecule type: mRNA  
A:Residues: 1-161 <GOP>  
A:Cross-references: EMBL:X54076; NID:g22338; PIDN:CAA38013.1; PID:g22339  
C:Genetics:  
A:Gene: hsp18-9  
A:Superfamily: alpha-crystallin  
C:Keywords: heat shock; stress-induced protein

Query Match 72.9%; Score 35; DB 2; Length 161;  
Best Local Similarity 85.7%; Pred. No. 42;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 GGEKAGG 8  
||| |||  
Db 30 GGSKAGG 36

RESULT 14

H69074

formylmethanofuran dehydrogenase (EC 1.2.99.5) (tungsten) chain C - Methanobacterium ther  
C:Species: Methanobacterium thermoautotrophicum  
C>Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 24-Sep-1999  
C:Accession: H69074  
R:Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.; Dubois, J.; Aldredge, T.; I  
Qiu, D.; Spadafora, R.; Vicaire, R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiواني, N.;  
ki, S.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.

J. Bacteriol. 179, 7135-7155, 1997

A:Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta H: functi  
A:Reference number: A69000; MUID:98037514; PMID:9371463  
A:Accession: H69074  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA

A;Residues: 1-270 <MTH>  
A;Cross-references: GB:AE000916; GB:AE000666; NID:g2622674; PIDN:AA86032.1; PID:g262268  
A;Experimental source: strain Delta H  
C;Genetics:  
A;Gene: MTH1558  
C;Superfamily: formylmethanofuran dehydrogenase chain C  
C;Keywords: iron-sulfur protein; metalloprotein; oxidoreductase; tungsten

Query Match 72.9%; Score 35; DB 2; Length 270;  
Best Local Similarity 75.0%; Pred. No. 70;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 IGGEKAGG 8  
:|||||  
Db 200 VGEWAGG 207

# RESULT 15

S57457  
formylmethanofuran dehydrogenase (EC 1.2.99.5) (tungsten) chain C - Methanobacterium the  
C;Species: Methanobacterium thermoautotrophicum  
C;Date: 10-Oct-1995 #sequence\_revision 17-Nov-1995 #text\_change 20-Jun-2000  
C;Accession: S63546; S57457  
R;Hochheimer, A.; Schmitz, R.A.; Thauer, R.K.; Hedderich, R.  
Eur. J. Biochem. 234, 910-920, 1995  
A;Title: The tungsten formylmethanofuran dehydrogenase from Methanobacterium thermoautot  
A;Reference number: S63519; MUID:96163477; PMID:8575452  
A;Accession: S63546  
A;Molecule type: DNA  
A;Residues: 1-270 <HOC>  
A;Cross-references: EMBL:X87970; NID:g1890205; PIDN:CAA61214.1; PID:g1890211  
A;Experimental source: strain Marburg, DSM 2133  
C;Genetics:  
A;Gene: fwdC

C;Complex: heterotetramer of A (see PIR:S57456), B (see PIR:S57458), C, and D (see PIR:S  
C;Superfamily: formylmethanofuran dehydrogenase chain C  
C;Keywords: heterotetramer; iron-sulfur protein; metalloprotein; oxidoreductase; tungste

Query Match 72.9%; Score 35; DB 2; Length 270;  
Best Local Similarity 75.0%; Pred. No. 70;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 IGGEKAGG 8  
:|||||  
Db 200 VGEWAGG 207

Search completed: September 18, 2004, 04:30:51  
Job time : 11.9412 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 18, 2004, 03:48:54 ; Search time 6.70588 Seconds  
(without alignments)  
69.884 Million cell updates/sec

Title: US-10-615-959-41  
Perfect score: 48  
Sequence: 1 IGGKAGGF 9

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_42.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	48	100.0	1466	1 CA13_HUMAN	P02461 homo sapien
2	44	91.7	636	1 CA13_RAT	P13941 rattus norv
3	44	91.7	1464	1 CA13_MOUSE	P08121 mus musculu
4	36	75.0	268	1 XH21_ECOLI	P39371 escherichia
5	36	75.0	629	1 K2C3_HUMAN	P12035 homo sapien
6	35	72.9	161	1 HS21_MAIZE	P24631 zea mays (m
7	35	72.9	270	1 FWDC_METTH	O27600 methanobact
8	35	72.9	270	1 FWDC_METTM	O59579 methanobact
9	35	72.9	382	1 OVEA_SCHPO	O09670 schizosacch
10	35	72.9	468	1 IL9R_MOUSE	Q01114 mus musculu
11	35	72.9	559	1 MERA_ALCSP	P94188 alcaligenes
12	35	72.9	1009	1 POL_SIVGB	P22382 simian immu
13	34	70.8	165	1 RL12_SCHPO	O75000 schizosacch
14	34	70.8	240	1 YGL2_STRVR	P19435 streptomyc
15	34	70.8	561	1 EST4_RAT	O64573 rattus norv
16	34	70.8	561	1 EST5_RAT	O63010 rattus norv
17	34	70.8	565	1 ES10_RAT	P16303 rattus norv
18	34	70.8	1137	1 RIK1_HSV11	P08543 herpes simp
19	33	68.8	50	1 YK05_CAEEL	P34300 caenorhabdi
20	33	68.8	265	1 VNST_CVMJH	P26625 murine coro
21	33	68.8	267	1 MTRC_METAC	O8TU01 methanosarc
22	33	68.8	312	1 YD63_THEAC	Q9H166 thermoplasm
23	33	68.8	374	1 TRMU_VIBPA	O8Wq19 vibrio para
24	33	68.8	374	1 TRMU_VIBVU	O8Wqj6 vibrio vuln
25	33	68.8	394	1 ASSY_METWA	O8G0U5 methanosarc
26	33	68.8	428	1 VG4_BP122	P15420 bacterioph
27	33	68.8	1145	1 DP2L_METAC	O8thg5 methanosarc
28	32	66.7	97	1 RL28_RICCN	Q92Jd1 rickettsia
29	32	66.7	97	1 RL28_RICPR	O9ze48 rickettsia
30	32	66.7	155	1 DHB_CRAPL	P22239 craterostig
31	32	66.7	173	1 SSB_AGR75	O8uf87 agrobacteri
32	32	66.7	214	1 SMP_ECOLI	P18838 escherichia
33	32	66.7	236	1 YEW3_SCHPO	O14141 schizosacch

34 32 66.7 270 1 FWDC\_METWO 074031 methanobact  
35 32 66.7 275 1 YX21\_CAUCR 09a313 caulobacter  
36 32 66.7 287 1 RL6\_HUMAN 002878 homo sapien  
37 32 66.7 287 1 RL6\_MOUSE 047911 mus musculu  
38 32 66.7 296 1 RL6\_RAT 021533 rattus norv  
39 32 66.7 302 1 HSLO\_NEIMA 09jwc8 neisseria m  
40 32 66.7 302 1 HSLO\_NEIMB 09jxx1 neisseria m  
41 32 66.7 339 1 FABH\_STRLI 09f6d4 streptomyce  
42 32 66.7 384 1 YH74\_VIBCH 09k669 vibrio chol  
43 32 66.7 484 1 PGD\_HAEIN 043774 haemophilus  
44 32 66.7 528 1 C318\_DROME 09vq35 drosophila  
45 32 66.7 592 1 RB56\_HUMAN 092804 homo sapien

## ALIGNMENTS

RESULT 1  
CAL3\_HUMAN STANDARD; PRT; 1466 AA.  
ID CA13\_HUMAN Q15112;  
AC P02461; Q15112;  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 01-JAN-1990 (Rel. 13, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Collagen alpha 1(III) chain precursor.  
GN COL3A1.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Skin fibroblast;  
RX MEDLINE=89350838; PubMed=2764886;  
RA Ala-Kokko L., Kontusaari S., Baldwin C.T., Kuivaniemi H.,  
RA Prockop D.J.;  
RT "Structure of cDNA clones coding for the entire prepro alpha 1 (III)  
chain of human type III procollagen. Differences in protein structure  
from type I procollagen and conservation of codon preferences.";  
RL Biochem. J. 260:509-516(1989).  
RN [2]  
RP SEQUENCE OF 149-1225 FROM N.A.  
RX MEDLINE=89386015; PubMed=2780304;  
RA Janeczko R.A., Ramirez F.;  
RT "Nucleotide and amino acid sequences of the entire human alpha 1  
(III) collagen.";  
RL Nucleic Acids Res. 17:6742-6742(1989).  
RN [3]  
RP SEQUENCE OF 168-398.  
RX MEDLINE=77134724; PubMed=557335;  
RA Seyer J.M., Kang A.H.;  
RT "Covalent structure of collagen: amino acid sequence of cyanogen  
bromide peptides from the amino-terminal segment of type III collagen  
of human liver.";  
RL Biochemistry 16:1158-1164(1977).  
RN [4]  
RP REVISIONS.  
RA Seyer J.M.;  
RL Submitted (DEC-1977) to the PIR data bank.  
RN [5]  
RP SEQUENCE OF 399-727.  
RX MEDLINE=79000343; PubMed=687591;  
RA Seyer J.M., Kang A.H.;  
RT "Covalent structure of collagen: amino acid sequence of five  
consecutive CNBr peptides from type III collagen of human liver.";  
RL Biochemistry 17:3404-3411(1978).  
RN [6]  
RP SEQUENCE OF 728-964.  
RX MEDLINE=80198282; PubMed=6246925;  
RA Seyer J.M., Mainardi C., Kang A.H.;  
RT "Covalent structure of collagen: amino acid sequence of alpha 1  
(III)-CB5 from type III collagen of human liver.";  
RL Biochemistry 19:1583-1589(1980).

RN [7] SEQUENCE OF 950-1466 FROM N.A.  
RP MEDLINE=88189827; PubMed=3357782;  
RA Markoo B.S., Dalglish R.;  
RT "Human pro alpha 1(III) collagen: cDNA sequence for the 3' end.";  
RL Nucleic Acids Res. 16:2337-2337(1988).  
RN [8] REVISION TO 1184.  
RP MEDLINE=89098346; PubMed=3211760;  
RA Molyneux K., Dalglish R.;  
RT "Molecular cloning and carboxyl-propeptide analysis of human type III procollagen";  
RL Nucleic Acids Res. 12:9383-9394(1984).  
RN [10] SEQUENCE OF 965-1200.  
RP MEDLINE=81208139; PubMed=7016180;  
RA Seyer J.M., Kang A.H.;  
RT "Covalent structure of collagen: amino acid sequence of alpha 1(III)-C99 from type III collagen of human liver.";  
RL Biochemistry 20:2621-2627(1981).  
RN [11] SEQUENCE OF 1176-1466 FROM N.A.  
RP MEDLINE=85157600; PubMed=2579949;  
RA Chu M.-L., Weil D., de Wet W.J., Bernard M.P., Sippola M., Ramirez F.;  
RT "Isolation of cDNA and genomic clones encoding human pro-alpha 1 (III) collagen. Partial characterization of the 3' end region of the gene.";  
RL J. Biol. Chem. 260:4357-4363(1985).  
RN [12] SEQUENCE OF 1161-1200 FROM N.A.  
RP MEDLINE=86187804; PubMed=3754462;  
RA Miskulin M., Dalglish R., Kluge-Beckerman B., Renard S.I., Tolstoshev P., Brantly R., Crystal R.G.;  
RT "Human type III collagen gene expression is coordinately modulated with the type I collagen genes during fibroblast growth.";  
RL Biochemistry 25:1408-1413(1986).  
RN [13] SEQUENCE OF 1-170 FROM N.A.  
RP TISSUE=Placenta;  
RX MEDLINE=88303360; PubMed=3405773;  
RA Tomar D., Ricca G., de Crombrughe B.;  
RT "Nucleotide sequence of a cDNA coding for the amino-terminal region of human prepro alpha 1(III) collagen.";  
RL Nucleic Acids Res. 16:7201-7201(1988).  
RN [14] SEQUENCE OF 1-176 FROM N.A.  
RP MEDLINE=89378752; PubMed=2777083;  
RA Benson-Chanda V., Su M.W., Weil D., Chu M.-L., Ramirez F.;  
RT "Cloning and analysis of the 5' portion of the human type-III procollagen gene (COL3A1).";  
RL Gene 78:255-265(1989).  
RN [15] REVIEW ON VARIANTS.  
RP MEDLINE=97255959; PubMed=9101290;  
RA Kuivaniemi H., Tromp G., Prockop D.J.;  
RT "Mutations in fibrillar collagens (types I, II, III, and XI), fibrillar-associated collagen (type IX), and network-forming collagen (type X) cause a spectrum of diseases of bone, cartilage, and blood vessels.";  
RL Hum. Mutat. 9:300-315(1997).  
RN [16] VARIANT AORTIC ANEURYSM ARG-303, AND VARIANT THR-668.  
RX MEDLINE=93293988; PubMed=8514866;  
RA Tromp G., Wu Y., Prockop D.J., Madhathari S.L., Kleinert C., Earley J.J., Zhuang J., Nooragaard O., Darling R.C., Abbott W.M., Cole C.W., Jaakkola P., Rytynen M., Pearce W.H., Yao J.S.T., Majamaa K., Smullen S.N., Gatalica Z., Ferrell R.E., Jimenez S.A.;  
RT "Sequencing of cDNA from 50 unrelated patients reveals that mutations in the triple-helical domain of type III procollagen are an infrequent cause of aortic aneurysms.";  
RL J. Clin. Invest. 91:2539-2545(1993).  
RN [17] VARIANT THR-698.  
RP MEDLINE=91045136; PubMed=2235526;  
RA Zafarullah K., Kleinert C., Tromp G., Kuivaniemi H., Kontusaari S., Wu Y., Ganguly A., Prockop D.J.;  
RT "G to A polymorphism in exon 31 of the COL3A1 gene.";  
RL Nucleic Acids Res. 18:6180-6180(1990).  
RN [18] VARIANT AORTIC ANEURYSM ARG-786.  
RX MEDLINE=91056145; PubMed=2243125;  
RA Kontusaari S., Tromp G., Kuivaniemi H., Romanic A.M., Prockop D.J.;  
RT "A mutation in the gene for type III procollagen (COL3A1) in a family with aortic aneurysms.";  
RL J. Clin. Invest. 86:1465-1473(1990).  
RN [19] VARIANT EDS-IV ARG-828.  
RX MEDLINE=94016385; PubMed=8411057;  
RA Richards A.J., Narcisi P., Lloyd J.C., Prockop D.J.;  
RT "The substitution of glycine 661 by arginine in type III collagen produces mutant molecules with different thermal stabilities and causes Ehlers-Danlos syndrome type IV.";  
RL J. Med. Genet. 30:690-693(1993).  
RN [20] VARIANT EDS-IV SER-957.  
RX MEDLINE=89109135; PubMed=2492273;  
RA Tromp G., Kuivaniemi H., Shikata H., Prockop D.J.;  
RT "A single base mutation that substitutes serine for glycine 790 of the alpha 1 (III) chain of type III procollagen exposes an arginine and causes Ehlers-Danlos syndrome IV.";  
RL J. Biol. Chem. 264:1349-1352(1989).  
RN [21] VARIANT EDS-IV VAL-960.  
RX MEDLINE=95268429; PubMed=7749417;  
RA Tromp G., de Paeppe A., Nuytink L., Madhathari S.L., Kuivaniemi H.;  
RT "Substitution of valine for glycine 793 in type III procollagen in Ehlers-Danlos syndrome type IV.";  
RL Hum. Mutat. 5:179-181(1995).  
RN [22] VARIANT EDS-IV GLU-1014.  
RX MEDLINE=92316511; PubMed=1352273;  
RA Richards A.J., Ward P.N., Narcisi P., Nicholls A.C., Lloyd J.C., Pope F.M.;  
RT "A single base mutation in the gene for type III collagen (COL3A1) converts glycine 847 to glutamic acid in a family with Ehlers-Danlos syndrome type IV. An unaffected family member is mosaic for the mutation.";  
RL Hum. Genet. 89:414-418(1992).  
RN [23] VARIANT EDS-IV ASP-1050.  
RX MEDLINE=90037070; PubMed=2808425;  
RA Tromp G., Kuivaniemi H., Stolle C.A., Pope F.M., Prockop D.J.;  
RT "Single base mutation in the type III procollagen gene that converts the codon for glycine 883 to aspartate in a mild variant of Ehlers-Danlos syndrome IV.";  
RL J. Biol. Chem. 264:19313-19317(1989).  
RN [24] VARIANT EDS-IV VAL-1077.  
RX MEDLINE=91374480; PubMed=1895316;  
RA Richards A.J., Lloyd J.C., Ward P.N., de Paeppe A., Narcisi P., Pope F.M.;  
RT "Characterisation of a glycine to valine substitution at amino acid position 910 of the triple helical region of type III collagen in a patient with Ehlers-Danlos syndrome type IV.";  
RL J. Med. Genet. 28:458-463(1991).  
RN [25] VARIANT EDS-IV GLU-1173.  
RX MEDLINE=93022543; PubMed=1357232;  
RA Johnson P.H., Richards A.J., Pope F.M., Hopkinson D.A.;

```
Query Match 100.0%; Score 48; DB 1; Length 1466;
Best Local Similarity 100.0%; Pred. No. 0.68;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IGGEKAGGF 9
    |||||
DB 1208 IGGEKAGGF 1216

RESULT 2
ID CAL3_RAT STANDARD; PRT; 636 AA.
AC P13941; 070604;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Collagen alpha 1(III) chain (Fragment).
GN COL3A1.

OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94114571; PubMed=8286415;
RA Glumoff V., Maekela J.K., Vuorio E.;
RT "Cloning of cDNA for rat alpha 1(III) collagen mRNA. Different
RT expression patterns of type I and type III collagen and fibronectin
RT genes in experimental granulation tissue.";
RL Biochim. Biophys. Acta 1217:41-48(1994).
RN [2]
RP SEQUENCE OF 73-636 FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Fibroblast;
RA Wurtz T., Ellerstrom C., Lundmark C., Christerson C.;
RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 308-482 FROM N.A.
RX MEDLINE=88296083; PubMed=2456904;
RA Frankel F.R., Hsu C.-Y.J., Meyers J.C., Lin E., Lyttle C.R.,
RA Komm B., Mohr K.;
RT "Regulation of alpha 2(I), alpha 1(III), and alpha 2(V) collagen
RT mRNAs by estradiol in the immature rat uterus.";
RL DNA 7:347-354(1988).
CC -!- FUNCTION: Collagen type III occurs in most soft connective tissues
CC along with type I collagen.
CC -!- SUBUNIT: Trimers of identical alpha 1(III) chains. The chains are
CC linked to each other by interchain disulfide bonds. Trimers are
CC also cross-linked via hydroxylsines.
CC -!- PTM: Prolines at the third position of the tripeptide repeating
CC unit (G-X-Y) are hydroxylated in some or all of the chains.
CC
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement. (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch.)
CC
CC EMBL; X70369; CAA49832.1; -
CC DR EMBL; AJ005395; CAA06510.1; -
CC DR EMBL; M21354; AAA0942.1; -
CC DR FIR; S41067; S41067.
CC DR InterPro; IPR008161; Clg_helix.
CC DR InterPro; IPR008160; Collagen.
CC DR InterPro; IPR000895; Fib_collagen_C.
CC DR InterPro; IPR002181; Fibrogen_C.
CC DR InterPro; IPR001007; WVF_C.
CC Pfam; PF01410; COLFI; 1.
CC DR Pfam; PF01391; Collagen; 6.
CC DR ProDom; PD000007; Clg_helix; 1.
CC DR ProDom; PD002078; Fib_collagen_C; 1.
```

```
SMART; SM00038; COLFI; 1.
DR PROSITE; PS01208; WVF_C_1; PARTIAL.
KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW Collagen; Glycoprotein.
FT NON_TER 1
FT CHAIN <1 375 COLLAGEN ALPHA 1(III) CHAIN.
FT PROPEP 376 636 CARBOXYL-TERMINAL PROPEPTIDE.
FT DOMAIN <1 368 TRIPLE-HELICAL REGION.
FT DISULFID 369 636 NONHELICAL REGION (C-TERMINAL).
FT DISULFID 368 368 INTERCHAIN (BY SIMILARITY).
FT DISULFID 369 369 INTERCHAIN (BY SIMILARITY).
FT CONFLICT 340 340 N -> D (IN REF. 2).
FT CONFLICT 429 429 A -> G (IN REF. 2).
SQ SEQUENCE 636 AA; 62332 MW; 61A48159F01D01EE CRC64;

Query Match 91.7%; Score 44; DB 1; Length 636;
Best Local Similarity 77.8%; Pred. No. 1.6;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 IGGEKAGGF 9
    :|||:
DB 378 VGGEKSGGF 386

RESULT 3
CAL3_MOUSE STANDARD; PRT; 1464 AA.
ID CAL3_MOUSE
AC P08121; Q61429; Q9CRN7;
DT 01-AUG-1988 (Rel. 08, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Collagen alpha 1(III) chain precursor.
GN COL3A1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=C57BL/6 X DBA; TISSUE=Embryo;
RX MEDLINE=95011609; PubMed=7926795;
RA Tomlan D., de Crombrughe B.;
RT "The mouse type-III procollagen-encoding gene: genomic cloning and
RT complete DNA sequence.";
RL Gene 147:161-168(1994).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=C57BL/6; TISSUE=Brain;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.P., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udelsin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettman M., Maman A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley A.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinaki M.I., Skalska U., Smalish D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [3]
RP SEQUENCE OF 1-488 FROM N.A.
RX MEDLINE=88167858; PubMed=3443309;
RA Wood L., Theriault N., Vogeli G.;
```

"Complete nucleotide sequence of the N-terminal domains of the murine alpha-1 type-III collagen chain.";  
 Gene 61:225-230(1987).  
 [4]

SEQUENCE OF 1-28 FROM N.A.  
 MEDLINE=95131189; PubMed=3972847;  
 Liaw G., Mudryj M., de Crombrughe B.;  
 "Identification of the promoter and first exon of the mouse alpha 1  
 (III) collagen gene.";  
 J. Biol. Chem. 260:3773-3777(1985).  
 [5]

SEQUENCE OF 810-1464 FROM N.A.  
 STRAIN=C57BL/6J; TISSUE=Embryonic head;  
 MEDLINE=21085660; PubMed=11217851;  
 Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
 Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,  
 Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,  
 Saito T., Okazaki Y., Gojohori T., Bono H., Kasukawa T., Saito R.,  
 Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
 Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,  
 Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,  
 Schiraldi L.M., Staudli F., Suzuki R., Tomita M., Wagner L., Washio T.,  
 Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,  
 Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,  
 Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
 Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,  
 Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,  
 Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
 Suzuki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,  
 Hayashizaki Y., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,  
 Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,  
 Hayashizaki Y.;  
 "Functional annotation of a full-length mouse cDNA collection.";  
 Nature 409:685-690(2001).  
 [6]

SEQUENCE OF 1442-1464 FROM N.A.  
 STRAIN=C57BL/6;  
 MEDLINE=91274355; PubMed=2054384;  
 Metsaeranta M., Toman D., de Crombrughe B., Vuorio E.;  
 "Specific hybridization probes for mouse type I, II, III and IX  
 collagen mRNAs";  
 Biochim. Biophys. Acta 1089:241-243(1991).  
 -!- FUNCTION: Collagen type III occurs in most soft connective tissues  
 along with type I collagen.  
 -!- SUBUNIT: Trimers of identical alpha 1(III) chains. The chains are  
 linked to each other by interchain disulfide bonds. Trimers are  
 also cross-linked via hydroxylsines.  
 -!- PTM: Proline residues at the third position of the tripeptide  
 repeating unit (G-X-Y) are hydroxylated in some or all of the  
 chains.  
 -!- PTM: O-linked glycan consists of a Glc-Gal disaccharide bound to  
 the oxygen atom of a post-translationally added hydroxyl group (By  
 similarity).  
 -!- SIMILARITY: Contains 1 WFCC domain.

-----  
 This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 -----

EMBL; X52046; CAA36279.1; -;  
 EMBL; BC043089; AAH43089.1; -;  
 EMBL; BC058724; AAH58724.1; -;  
 EMBL; M18933; AAA37338.1; -;  
 EMBL; K03037; -; NOT ANNOTATED CDS.  
 EMBL; AK019448; BAB31724.1; -;  
 EMBL; X57983; CAA41048.1; -;  
 EMBL; A27353; A27353.  
 PIR; S59856; S59856.  
 MGD; MGI:88453; Col13a1.

DR InterPro; IPR008161; Clg\_helix.  
 DR InterPro; IPR008160; Collagen.  
 DR InterPro; IPR000885; Fib\_collagen\_C.  
 DR InterPro; IPR002181; Fibrinogen\_C.  
 DR InterPro; IPR001007; VWFC\_C.  
 DR Pfam; PF01410; COLFI; 1.  
 DR Pfam; PF01391; Collagen; 18.  
 DR ProDom; PD000007; Clg\_helix; 1.  
 DR ProDom; PD002078; Fib\_collagen\_C; 1.  
 DR SMART; SM00038; COLFI; 1.  
 DR SMART; SM00214; VWFC; 1.  
 DR PROSITE; PS01208; VWFC\_1; 1.  
 DR PROSITE; PS0184; VWFC\_2; 1.  
 KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;  
 Glycoprotein; Collagen; Signal.  
 FT SIGNAL 1 23 BY SIMILARITY.  
 FT CHAIN 24 154 AMINO-TERMINAL PROPEPTIDE.  
 FT CHAIN 155 1203 COLLAGEN ALPHA 1(III) CHAIN.  
 FT PROPEP 1204 1464 CARBOXYL-TERMINAL PROPEPTIDE.  
 FT DOMAIN 31 90 VWFC.  
 FT DOMAIN 155 169 NONHELICAL REGION (N-TERMINAL).  
 FT DOMAIN 170 1195 TRIPLE-HELICAL REGION.  
 FT DOMAIN 1196 1464 NONHELICAL REGION (C-TERMINAL).  
 FT CARBOHYD 262 262 O-LINKED (GAL. . .) (BY SIMILARITY).  
 FT MOD\_RES 262 262 HYDROXYLATION (BY SIMILARITY).  
 FT MOD\_RES 283 283 HYDROXYLATION (BY SIMILARITY).  
 FT MOD\_RES 859 859 HYDROXYLATION (BY SIMILARITY).  
 FT MOD\_RES 976 976 HYDROXYLATION (BY SIMILARITY).  
 FT MOD\_RES 1093 1093 HYDROXYLATION (BY SIMILARITY).  
 FT MOD\_RES 1105 1105 HYDROXYLATION (BY SIMILARITY).  
 FT DISULFID 1195 1195 INTERCHAIN (BY SIMILARITY).  
 FT DISULFID 1196 1196 INTERCHAIN (BY SIMILARITY).  
 SQ SEQUENCE 1464 AA; 138944 MW; 2104EC27A886090B CRC64;

Query Match 91.7%; Score 44; DB 1; Length 1464;  
 Best Local Similarity 77.8%; Pred. No. 3.7;  
 Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 IGGEKAGGF 9  
 :||||:|  
 Db 1206 VGGKSGGF 1214

## RESULT 4

YHHT\_ECOLI STANDARD; PRT; 368 AA.  
 ID YHHT\_ECOLI  
 AC P39371;  
 DT 01-FEB-1995 (Rel. 31, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Hypothetical protein yHHT precursor.  
 GN YHHT OR B4310.  
 OS Escherichia coli.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Escherichia.  
 OX NCBI\_TaxID=562;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12 / MG1655;  
 RX MEDLINE=95334362; PubMed=7610040;  
 RA Burland V.D., Plunkett G. III, Sofia H.J., Daniels D.L.,  
 RA Blattner F.R.;  
 RT "Analysis of the Escherichia coli genome VI: DNA sequence of the  
 RT region from 92.8 through 100 minutes.";  
 RL Nucleic Acids Res. 23:2105-2119(1995).  
 CC -!- SIMILARITY: Contains 7 Keich repeats.  
 CC -!- SIMILARITY: STRONG, TO H.INFLUENZAE H10148.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 -----

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; U14003; AAA97206.1; ALT\_INIT.  
 CC EMBL; A5000501; AAC77266.1; ALT\_INIT.  
 CC EcoGene; EGI2562; YJHT.  
 CC InterPro; IPR006652; Kelch\_rep.  
 CC DR Pfam; PF01344; Kelch; 2.  
 CC KW Hypothetical protein; Kelch repeat; Repeat; Signal; Complete proteome.  
 CC FT SIGNAL 1 19 POTENTIAL.  
 CC FT CHAIN 20 368 HYPOTHETICAL PROTEIN YJHT.  
 CC FT REPEAT 40 84 KELCH 1.  
 CC FT REPEAT 86 137 KELCH 2.  
 CC FT REPEAT 139 173 KELCH 3.  
 CC FT REPEAT 174 219 KELCH 4.  
 CC FT REPEAT 222 265 KELCH 5.  
 CC FT REPEAT 287 336 KELCH 6.  
 CC FT REPEAT 338 367 KELCH 7.  
 CC SQ SEQUENCE 368 AA; 39572 MW; 1194F392C51EA204 CRC64;

Query Match 75.0%; Score 36; DB 1; Length 368;  
 Best Local Similarity 87.5%; Pred. No. 28;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 IGGEKAGG 8  
 |||||  
 Db 341 IGGETAGG 348

RESULT 5  
 ID K2C3 HUMAN STANDARD; PRT; 629 AA.  
 AC P12035;  
 DT 01-OCT-1989 (Rel. 12, Created)  
 DT 01-OCT-1989 (Rel. 12, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE Keratin, type II cytoskeletal 3 (Cytokeratin 3) (K3) (CK3) (65 kDa  
 DE cytokeratin).  
 GN KRT3.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 ON NCBI\_TaxID=9606;  
 RX [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=87254239; PubMed=2439698;  
 RA Klinge E.M., Sylvestre Y.R., Freedberg I.M., Blumenberg M.;  
 RT "Evolution of keratin genes: different protein domains evolve by  
 RT different pathways."  
 RL J. Mol. Evol. 24:319-329(1987).  
 RN [2]

RP VARIANT MCD LYS-509.  
 RX MEDLINE=97315826; PubMed=9171831;  
 RA Irvine A.D., Corden L.D., Swenson O., Swenson B., Moore J.E.,  
 RA Frazer D.G., Smith F.J.D., Knowlton R.G., Christophers E.,  
 RA Rochels R., Witto J., McLean W.H.I.;  
 RT "Mutations in cornea-specific keratin K3 or K12 genes cause  
 RT Meesmann's corneal dystrophy."  
 RL Nat. Genet. 16:184-187(1997).  
 CC -1- SUBUNIT: Heterotrimer of two type I and two type II keratins.  
 CC Keratin 3 associates with keratin 12.  
 CC -1- TISSUE SPECIFICITY: Cornea specific.  
 CC -1- DISEASE: Defects in KRT3 are a cause of Meesmann corneal dystrophy  
 CC (MCD) [MIM:122100]. MCD is an autosomal dominant disease that  
 CC causes fragility of the anterior corneal epithelium. Patients are  
 CC usually asymptomatic until adulthood when rupture of the corneal  
 CC microcysts may cause erosions, producing clinical symptoms such as  
 CC photophobia, contact lens intolerance and intermittent diminution  
 CC of visual acuity. Rarely, subepithelial scarring causes irregular  
 CC corneal astigmatism and permanent visual impairment. Histological  
 CC examination shows a disorganized and thickened epithelium with  
 CC widespread cytoplasmic vacuolation and numerous small, round,  
 CC debris-laden intraepithelial cysts.

CC -1- MISCELLANEOUS: There are two types of cytoskeletal and  
 CC microfibrillar keratin: I (acidic; 40-55 kDa) [K9 to K20] and II  
 CC (neutral to basic; 56-70 kDa) [K1 to K8].  
 CC -1- SIMILARITY: Belongs to the intermediate filament family.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; X05418; CAA28991.1; -  
 CC EMBL; X05419; CAA28992.1; ALT\_SEQ.  
 CC EMBL; X05420; CAA28993.1; ALT\_SEQ.  
 CC EMBL; X05420; CAA28994.1; ALT\_SEQ.  
 CC EMBL; X05420; CAA28995.1; ALT\_SEQ.  
 CC EMBL; X05421; CAA28996.1; ALT\_SEQ.  
 CC Genew; HGNC:6440; KRT3.  
 CC MIM; 148043; -  
 CC MIM; 122100; -  
 CC GO; GO:0005882; C:intermediate filament; NAS.  
 CC GO; GO:0005198; F:structural molecule activity; NAS.  
 CC GO; GO:0007010; P:cytoskeleton organization and biogenesis; NAS.  
 CC InterPro; IPR001664; IF.  
 CC InterPro; IPR002957; Keratin\_I.  
 CC InterPro; IPR003054; Keratin\_II.  
 CC Pfam; PF00038; Filament; 1.  
 CC PRINTS; PR01248; TYPE1KERATIN.  
 CC PRINTS; PR01276; TYPE2KERATIN.  
 CC PROSITE; PS00226; IF; 1.  
 CC Intermediate filament; Coiled coil; Keratin; Phosphorylation;  
 KW Disease mutation.  
 FT DOMAIN 1 197 HEAD.  
 FT DOMAIN 198 509 ROD.  
 FT DOMAIN 510 629 TAIL.  
 FT DOMAIN 198 233 COIL 1A.  
 FT DOMAIN 234 254 LINKER 1.  
 FT DOMAIN 255 346 COIL 1B.  
 FT DOMAIN 347 370 LINKER 12.  
 FT DOMAIN 371 509 COIL 2.  
 FT MOD\_RES 56 56 PHOSPHORYLATION (BY SIMILARITY).  
 FT VARIANT 509 509 E -> K (in MCD).  
 FT /FTID=VAR\_003868.  
 SQ SEQUENCE 629 AA; 64511 MW; 2E748619A828BCD9 CRC64;

Query Match 75.0%; Score 36; DB 1; Length 629;  
 Best Local Similarity 75.0%; Pred. No. 47;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 GGKAGG 9  
 |||||  
 Db 73 GGSRAGG 80

RESULT 6  
 ID HS21 MAIZE STANDARD; PRT; 161 AA.  
 AC P24631;  
 DT 01-MAR-1992 (Rel. 21, Created)  
 DT 01-MAR-1992 (Rel. 21, Last sequence update)  
 DT 01-JUN-1994 (Rel. 29, Last annotation update)  
 DE 17.5 kDa class II heat shock protein.  
 OS Zea mays (Maize).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC PACAD clade; Panicoideae; Andropogoneae; Zea.  
 ON NCBI\_TaxID=4577;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Ohio 43; TISSUE=Radicle;  
 RX MEDLINE=91329703; PubMed=1714322;





CC -!- INDUCTION: By growth on tungsten or molybdenum under anaerobic  
 CC conditions.  
 CC -!- SIMILARITY: Belongs to the fwdC/fmcC family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; X87970; CAA61214.1; --  
 CC InterPro; IPR002489; DUF14.  
 CC Pfam; PF01493; GXGKG; 1.  
 CC Oxidoreductase; Tungsten; Mechanogenesis; Repeat.  
 CC X- [IL]-X- [IV]-X-G.  
 CC DOMAIN 80 213  
 CC FT REPEAT 80 92  
 CC FT REPEAT 99 111  
 CC FT REPEAT 118 130  
 CC FT REPEAT 144 156  
 CC FT REPEAT 163 175  
 CC FT REPEAT 182 194  
 CC FT REPEAT 201 213  
 CC SEQUENCE 270 AA; 28554 MW; E0A369D2ACFEC46F CRC64;  
 CC  
 CC Query Match 72.9%; Score 35; DB 1; Length 270;  
 CC Best Local Similarity 75.0%; Pred. No. 32;  
 CC Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
 CC  
 CC QY 1 IGGERAGG 8  
 CC :|||:|  
 CC Db 200 VGGWAGG 207  
 CC  
 CC RESULT 9  
 CC ID OYEA SCHPO STANDARD; PRT; 382 AA.  
 CC AC Q09670;  
 CC DT 01-NOV-1995 (Rel. 32, Created)  
 CC DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 CC DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 CC DE Putative NADPH dehydrogenase CSH10.04 (EC 1.6.99.1) (Old yellow enzyme  
 CC homolog).  
 CC GN SPAC5H10.04.  
 CC OS Schizosaccharomyces pombe (Fission yeast).  
 CC OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
 CC Schizosaccharomycetales; Schizosaccharomycetaceae;  
 CC Schizosaccharomycetes.  
 CC OX NCBI\_TaxID=4896;  
 CC RN [1]  
 CC RP SEQUENCE FROM N.A.  
 CC RC STRAIN=972;  
 CC RX MEDLINE=21848401; PubMed=11859360;  
 CC RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,  
 CC RA Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,  
 CC RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,  
 CC RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,  
 CC RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,  
 CC RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,  
 CC RA James K., Jones I., Jones M., Leather S., McDonald S., McLean J.,  
 CC RA Mooney P., Moule S., Mungall K., Murphy I., Niblett D., Odell C.,  
 CC RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,  
 CC RA Rutherford K., Rutter S., Saunders R., Seeger K., Sharp S.,  
 CC RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,  
 CC RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,  
 CC RA Woodward J., Volkart G., Aert R., Robben J., Grynoprez B.,  
 CC RA Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,  
 CC RA Gabel C., Fuchs M., Fritzc C., Holzer E., Moesli D., Hilbert H.,  
 CC RA Borzym K., Langer I., Beck A., Leirach H., Reinhardt R., Pohl T.M.,  
 CC RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,  
 CC RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,

RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,  
 RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,  
 RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,  
 RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,  
 RA Carrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,  
 RA Shpakowski G.V., Ussery D., Barrell B.G., Nurse P.;  
 RA "The genome sequence of Schizosaccharomyces pombe.";  
 RL Nature 415:871-880(2002).  
 CC -!- CATALYTIC ACTIVITY: NADPH + acceptor = NADP(+) + reduced acceptor.  
 CC -!- COFACTOR: FMN.  
 CC -!- SUBUNIT: Homodimer or heterodimer (By similarity).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; Z49811; CAA89954.1; --  
 CC PIR; S55482; S55482.  
 CC HSP; Q02899; IOYC.  
 CC GeneDB\_SPombe; SPAC5H10.04; --  
 CC InterPro; IPR001155; Oxidoreductase.  
 CC Pfam; PF00724; Oxidoreductase; FMN; 1.  
 CC Hypothetical protein; Oxidoreductase; NADP; Flavoprotein; FMN;  
 CC Multigene family.  
 CC SEQUENCE 382 AA; 43813 MW; 3F7939599CA167D1 CRC64;  
 CC  
 CC Query Match 72.9%; Score 35; DB 1; Length 382;  
 CC Best Local Similarity 85.7%; Pred. No. 44;  
 CC Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 CC  
 CC QY 3 GEKAGGF 9  
 CC :|||:|  
 CC Db 65 GEKSGGF 71  
 CC  
 CC RESULT 10  
 CC ID IL9R MOUSE STANDARD; PRT; 468 AA.  
 CC AC Q01114;  
 CC DT 01-APR-1993 (Rel. 25, Created)  
 CC DT 01-APR-1993 (Rel. 25, Last sequence update)  
 CC DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 CC DE Interleukin-9 receptor precursor (IL-9R).  
 CC GN IL9R.  
 CC OS Mus musculus (Mouse).  
 CC OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 CC OX NCBI\_TaxID=10090;  
 CC RN [1]  
 CC RP SEQUENCE FROM N.A.  
 CC RC TISSUE=T-cell;  
 CC RX MEDLINE=92302307; PubMed=1376929;  
 CC RA Renauld J.C., Druet C., Kermouni A., Houssiau F., Uytendhoe C.,  
 CC RA van Roost E., van Snick J.;  
 CC RT "Expression cloning of the murine and human interleukin 9 receptor  
 CC cDNAs".  
 CC RL Proc. Natl. Acad. Sci. U.S.A. 89:5690-5694(1992).  
 CC -!- FUNCTION: This is a receptor for interleukin-9.  
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein and secreted.  
 CC -!- SIMILARITY: Belongs to the type I cytokine family of receptors.  
 CC Subfamily 4.  
 CC -!- SIMILARITY: Contains 1 fibronectin type III domain.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

```
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M84746; AAA37871.1; -.
DR FIRM; A45268; A45268
DR MGD; MGI:96564; 139r.
DR InterPro; IPR002996; CRI1A.
DR InterPro; IPR008957; FN_III-like.
DR InterPro; IPR003531; Hemtopoptn S Fl.
DR PROSITE; PS01355; HEMATOPO REC S Fl.
KW Receptor; Transmembrane; Glycoprotein; signal; T-cell.
FT SIGNAL 1 37
FT CHAIN 38 468
FT DOMAIN 38 270
FT TRANSMEM 271 291
FT DOMAIN 292 468
FT DOMAIN 149 242
FT CARBOHYD 116 116
FT CARBOHYD 155 155
SQ SEQUENCE 468 AA; 52260 MW; BBE179FD72E29A5 CRC64;

Query Match 72.9%; Score 35; DB 1; Length 468;
Best Local Similarity 75.0%; Pred. No. 54;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 GGEKAGGF 9
Db 46 GGEKAGGF 53

RESULT 11
MERA ALCSP STANDARD; PRT; 559 AA.
AC P94188;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Mercuric reductase (EC 1.16.1.1) (Hg(II) reductase).
GN MERA.
OS Alkaligenes sp.
OG Plasmid IncH2 PMER610.
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Alkaligenaceae; Alkaligenes.
OX NCBI_TaxID=512;
RN [1]
RP SEQUENCE FROM N.A.
RA Nikiforov V., Yuriev O., Kholodii G., Minakhin L., Gorlenko Z.,
RA Kalyaeva E., Mindlin S.;
RA Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.
RL by a specialized system which includes mercuric reductase. Mera
CC protein is responsible for volatilizing mercury as Hg(0).
CC -!- FUNCTION: Resistance to Hg(2+) in bacteria appears to be governed
CC by a specialized system which includes mercuric reductase. Mera
CC -!- CATALYTIC ACTIVITY: Hg + NADP(+) + H(+) = Hg(2+) + NADPH.
CC -!- COFACTOR: Binds 1 FAD per subunit (By similarity).
CC -!- SUBUNIT: Homodimer (By similarity).
CC -!- MISCELLANEOUS: The active site is a redox-active disulfide bond.
CC -!- SIMILARITY: Belongs to class-I pyridine nucleotide-disulfide
CC oxidoreductase family.
CC -!- SIMILARITY: Contains 1 HMA domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Y08993; CAA70190.1; -.
DR HSSP; Q04656; LAW0.
DR InterPro; IPR001327; FAD_pyr_redox.
DR InterPro; IPR006121; HeavyMe transpt.
DR InterPro; IPR000815; Hg reductase.
DR InterPro; IPR006131; Metal_bind.
```

```
DR InterPro; IPR001100; Pyr_redox.
DR InterPro; IPR004099; pyr_redox_dim.
DR Pfam; PF00403; HMA_1.
DR Pfam; PF00070; pyr_redox; 1.
DR Pfam; PF02852; pyr_redox_dim; 1.
DR PRINTS; PR00368; FADPNR.
DR PRINTS; PR00945; HGRDTASE1.
DR PRINTS; PR00411; PNDRDTASE1.
DR ProDom; PD000139; FAD_pyr_redox; 1.
DR PROSITE; PS01047; HMA_1; 1.
DR PROSITE; PS00846; HMA_2; 1.
DR PROSITE; PS00076; PYRIDINE REDOX 1; 1.
KW Mercuric resistance; Oxidoreductase; Flavoprotein; FAD; NADP;
KW Mercury; Redox-active center; Metal-binding; Plasmid.
FT DOMAIN 1 65 HMA FAD (ADP PART) (BY SIMILARITY).
FT NP_BIND 126 134 REDOX-ACTIVE.
FT DISULFID 134 139 MERCURY (POTENTIAL).
FT METAL 556 556 MERCURY (POTENTIAL).
FT METAL 557 557 MERCURY (POTENTIAL).
SQ SEQUENCE 559 AA; 58019 MW; 0603DD6813CB519 CRC64;

Query Match 72.9%; Score 35; DB 1; Length 559;
Best Local Similarity 85.7%; Pred. No. 64;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 GGEKAGG 8
Db 87 GGEKAGG 93

RESULT 12
POL_SIVGB STANDARD; PRT; 1009 AA.
AC P22382;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE POL polyprotein [Contains: Protease (Retropepsin) (EC 3.4.23.-);
DE Reverse transcriptase (EC 2.7.7.49); Ribonuclease H (EC 3.1.26.4)].
GN POL.
OS Simian immunodeficiency virus (isolate GB1).
OC Viruses; Retrod viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11732;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=90015168; PubMed=2797181;
RA Tsujimoto H., Hasegawa A., Maki N., Fukasawa M., Miura T., Speidel S.,
RA Cooper R.W., Moriyama E.N., Gojobori T., Hayami M.;
RA "Sequence of a novel simian immunodeficiency virus from a wild-caught
RA African mandrill.";
RL Nature 341:539-541(1989).
CC -!- CATALYTIC ACTIVITY: Endonucleolytic cleavage to 5'-
CC phosphomonoester.
CC -!- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate = N diphosphate
CC + [DNA] (N).
CC -!- PTM: Cleavage sites that yield the mature proteins remain to be
CC determined.
CC -!- MISCELLANEOUS: THIS IS AN AFRICAN MANDRILL ISOLATE.
CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY A2; ALSO
CC KNOWN AS THE RETROPEPSIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M27470; AAB49569.1; -.
DR HSSP; P03366; 1HRH.
DR MEROPS; A02.003; -.
DR InterPro; IPR001969; Aspprotease_AS.
```

DR InterPro; IPR001037; Integrase C.  
 DR InterPro; IPR003308; Integrase Zn.  
 DR InterPro; IPR009007; Pept\_A\_acid.  
 DR InterPro; IPR001995; Peptidase\_A2.  
 DR InterPro; IPR002156; RNaseH.  
 DR InterPro; IPR001584; Rve.  
 DR InterPro; IPR000477; RVTe.  
 DR Pfam; PF00552; integrase.1.  
 DR Pfam; PF02022; integrase\_Zn; 1.  
 DR Pfam; PF00675; rnaseh; 1.  
 DR Pfam; PF00665; rve; 1.  
 DR Pfam; PF00077; rvp; 1.  
 DR Pfam; PF00078; rvt; 1.  
 DR PROSITE; PS00141; ASP PROTEASE; 1.  
 DR PROSITE; PS00175; ASP PROT RETROV; 1.  
 KW AIDS; Polyprotein; Hydrolase; Aspartyl protease; Endonuclease;  
 FT Nuclease; Transferase; RNA-directed DNA polymerase.  
 FT ACT\_SITE 86 86 BY SIMILARITY.  
 SQ SEQUENCE 1009 AA; 114954 MW; 37ED76A8ADE3821D CRC64;  
 Query Match 72.9%; Score 35; DB 1; Length 1009;  
 Best Local Similarity 75.0%; Pred. No. 1;le+02;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
 QY 2 GGEKAGGF 9  
 |||||  
 Db 52 GGEERGPF 59  
 RESULT 13  
 RL12 SCHPO STANDARD; PRT; 165 AA.  
 AC 075000;  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 25-DEC-1998 (Rel. 37, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE 60S ribosomal protein L12.  
 GN (RPL12A OR SPC33H12.04C) AND (RPL12B OR SPC16C4.13C).  
 OS Schizosaccharomyces pombe (Fission yeast).  
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
 OX NCBI\_TaxID=4896;  
 RN [1]  
 RP SEQUENCE FROM N.A. (RPL12A AND RPL12B).  
 RC STRAIN=972;  
 RX MEDLINE=21848401; PubMed=11859360;  
 RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,  
 RA Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,  
 RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,  
 RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,  
 RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,  
 RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,  
 RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,  
 RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,  
 RA Oliver K., O'Neill S., Pearson D., Quayl M.A., Rabinowitsch E.,  
 RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,  
 RA Skelton J., Simmonds M., Squares R., Squares K., Stevens K.,  
 RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,  
 RA Woodward J., Voiclaert G., Aert R., Robben J., Grymonprez B.,  
 RA Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,  
 RA Gabel C., Fuchs M., Fritz C., Holzer E., Woestl D., Hilbert H.,  
 RA Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,  
 RA Egar P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,  
 RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,  
 RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,  
 RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,  
 RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,  
 RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,  
 RA Cerrutti L., Lowe I., McCombie W.R., Paulsen I., Potashkin J.,  
 RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;  
 FT "The genome sequence of Schizosaccharomyces pombe."  
 RL Nature 415:871-880(2002).

CC -!- FUNCTION: This protein binds directly to 26S ribosomal RNA (By  
 CC similarity).  
 CC -!- MISCELLANEOUS: There are two genes for L12 in S.pombe.  
 CC -!- SIMILARITY: Belongs to the L11p family of ribosomal proteins.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC -----  
 CC EMBL; AL031824; CAA21221.1; -;  
 CC EMBL; AL031535; CAA20752.1; -;  
 DR PIR; T41103; T41103.  
 DR HSSP; P29395; 1MNS.  
 DR GeneDB\_Spombe; SPC33H12.04c; -;  
 DR GeneDB\_Spombe; SPC16C4.13c; -;  
 DR InterPro; IPR000911; Ribosomal\_L11.  
 DR Pfam; PF00298; Ribosomal\_L11; 1.  
 DR Pfam; PF03946; Ribosomal\_L11\_N; 1.  
 DR SMART; SM00649; RL11; 1.  
 DR PROSITE; PS00359; RIBOSOMAL\_L11; 1.  
 DR Ribosomal protein; RNA-binding; Multigene family.  
 SQ SEQUENCE 165 AA; 17666 MW; FC77B46CD08F005 CRC64;  
 Query Match 70.8%; Score 34; DB 1; Length 165;  
 Best Local Similarity 75.0%; Pred. No. 30;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 IGGEKAGG 8  
 :|||  
 Db 18 VGGEVAGG 25  
 RESULT 14  
 YGL2 STRVR  
 ID YGL2 STRVR STANDARD; PRT; 240 AA.  
 AC P19435;  
 DT 01-FEB-1991 (Rel. 17, Created)  
 DT 01-FEB-1991 (Rel. 17, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Hypothetical 23.9 kDa protein in glnII region (ORF2).  
 OS Streptomyces viridochromogenes.  
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 OC Streptomycineae; Streptomycetaceae; Streptomyces.  
 OX NCBI\_TaxID=1938;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ES2;  
 RX MEDLINE=90368592; PubMed=1975583;  
 RA Behrmann I., Hillemann D., Puehler A., Strauch E., Wohlleben W.;  
 RA "Overexpression of a Streptomyces viridochromogenes gene (glnII)  
 RT encoding a glutamine synthetase similar to those of eucaryotes confers  
 RT resistance against the antibiotic phosphinothricyl-alanyl-alanine";  
 RL J. Bacteriol. 172:5326-5334 (1990).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC -----  
 CC EMBL; X52842; CAA37027.1; -;  
 DR PIR; B36724; B36724.  
 KW Hypothetical protein.  
 SQ SEQUENCE 240 AA; 23914 MW; 09DBGBER681F7E2F CRC64;  
 Query Match 70.8%; Score 34; DB 1; Length 240;  
 Best Local Similarity 85.7%; Pred. No. 43;

Matches	6;	Conservative	1;	Mismatches	0;	Indels	0;	Gaps	0;
QY	2	GGEKAGG	8						
Db	118	GGEQAGG	124						
RESULT 15									
EST4 RAT									
ID	EST4 RAT	STANDARD;		PRT;	561 AA.				
AC	Q64573; Q62679;								
DT	01-NOV-1997 (Rel. 35, Created)								
DT	01-NOV-1997 (Rel. 35, Last sequence update)								
DT	28-FEB-2003 (Rel. 41, Last annotation update)								
DE	Liver carboxylesterase 4 precursor (EC 3.1.1.1) (Carboxylesterase ES-4)								
DE	(Microsomal palmitoyl-CoA hydrolase) (Kidney microsomal carboxylesterase).								
OS	Rattus norvegicus (Rat).								
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;								
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.								
OX	NCBI_TaxID=10116;								
RN	[1]								
RP	SEQUENCE FROM N.A.								
RC	STRAIN=Sprague-Dawley; TISSUE=Liver;								
RX	MEDLINE=96190723; PubMed=8611161;								
RA	Robbi M., van Schaftingen E., Beaufay H.;								
RT	"Cloning and sequencing of rat liver carboxylesterase ES-4 (microsomal palmitoyl-CoA hydrolase).";								
RL	Biochem. J. 313:821-826 (1996).								
RN	[2]								
RP	SEQUENCE FROM N.A.								
RC	STRAIN=Sprague-Dawley; TISSUE=Kidney;								
RX	MEDLINE=95050819; PubMed=7961958;								
RA	Yan B., Yang D., Brady M., Parkinson A.;								
RT	"Rat kidney carboxylesterase. Cloning, sequencing, cellular localization, and relationship to rat liver hydrolase.";								
RL	J. Biol. Chem. 269:29688-29696 (1994).								
CC	-1- FUNCTION: INVOLVED IN THE DETOXIFICATION OF XENOBIOTICS AND IN THE ACTIVATION OF ESTER AND AMIDE PRODRUGS. HYDROLYZES OMICRON-NITROPHENYL ACETATE AND ALPHA-NAPHTHYL ACETATE. IT ALSO HYDROLYZES ACETANILIDE AND, DISTINCTIVELY, PALMITOYL-COA.								
CC	-1- CATALYTIC ACTIVITY: A carboxylic ester + H(2)O = an alcohol + a carboxylic anion.								
CC	-1- SUBCELLULAR LOCATION: Microsomal membrane, lumen of endoplasmic reticulum.								
CC	-1- SIMILARITY: Belongs to the type-B carboxylesterase/lipase family.								
CC	-----								
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a> or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).								
CC	-----								
DR	EMBL; X81825; CAA57419.1; -								
DR	EMBL; U10697; AAB64638.1; -								
DR	PIR; S62788; S62788.								
DR	HSP; P21836; 1MAA.								
DR	InterPro; IPR002018; CarbesteraseB.								
DR	InterPro; IPR000886; ER_target_S.								
DR	InterPro; IPR000379; Ser_estrs.								
DR	Pfam; PF00135; Coesterase; 1.								
DR	PROSITE; PS00014; ER_TARGET; 1.								
DR	PROSITE; PS00122; CARBOXYLESTERASE_B_2; 1.								
DR	PROSITE; PS00941; CARBOXYLESTERASE_B_2; 1.								
KW	Glycoprotein; Hydrolase; Serine esterase; Endoplasmic reticulum; Signal; Multigene family.								
FT	SIGNAL	1	18						
FT	CHAIN	19	561						
FT	ACT_SITE	221	221						
FT	ACT_SITE	466	466						
FT	DISULFID	87	116						

FT	DISULFID	273	284	BY SIMILARITY.					
FT	SITE	558	561	PREVENT SECRETION FROM ER (POTENTIAL).					
FT	CARBOHYD	79	79	N-LINKED (GLCNAC. . .) (POTENTIAL).					
FT	CONFLICT	7	7	I -> F (IN REF. 2).					
FT	CONFLICT	60	60	L -> P (IN REF. 2).					
FT	CONFLICT	213	213	G -> A (IN REF. 2).					
FT	CONFLICT	253	253	P -> T (IN REF. 2).					
FT	CONFLICT	310	311	DN -> IT (IN REF. 2).					
FT	CONFLICT	342	342	T -> N (IN REF. 2).					
FT	CONFLICT	425	426	SI -> FY (IN REF. 2).					
FT	CONFLICT	509	509	G -> A (IN REF. 2).					
FT	CONFLICT	553	553	Q -> E (IN REF. 2).					
SQ	SEQUENCE	561 AA;	62274 MW;	EDF48F3309521C79	CRC64;				
Query Match 70.8%; Score 34; DB 1; Length 561;									
Best Local Similarity 77.8%; Pred. No. 98;									
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;									
QY	1	IGGEKAGGF	9						
Db	217	IFGESAGGF	225						
Search completed: September 18, 2004, 04:26:22									
Job time : 7.70588 secs									

FT DISULFID 273 284  
FT SITE 558 561  
FT CARBOHYD 79 79  
FT CONFLICT 7 7  
FT CONFLICT 60 60  
FT CONFLICT 213 213  
FT CONFLICT 253 253  
FT CONFLICT 310 311  
FT CONFLICT 342 342  
FT CONFLICT 425 426  
FT CONFLICT 509 509  
FT CONFLICT 553 553  
SQ SEQUENCE 561 AA; 62274 MW; EDF48F3309521C79 CRC64;

Query Match 70.8%; Score 34; DB 1; Length 561;  
Best Local Similarity 77.8%; Pred. No. 98;  
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 IGGEKAGGF 9  
Db 217 IFGESAGGF 225

Search completed: September 18, 2004, 04:26:22  
Job time : 7.70588 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 18, 2004, 04:00:08 ; Search time 31.9412 Seconds  
(without alignments)  
88.903 Million cell updates/sec

Title: US-10-615-959-41  
Perfect score: 48  
Sequence: 1 IGGEXAGGF 9

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SPTREMBL\_25:\*\*

- 1: sp\_archaea:\*\*
- 2: sp\_bacteria:\*\*
- 3: sp\_fungi:\*\*
- 4: sp\_human:\*\*
- 5: sp\_invertebrate:\*\*
- 6: sp\_mammal:\*\*
- 7: sp\_mhc:\*\*
- 8: sp\_organelle:\*\*
- 9: sp\_phage:\*\*
- 10: sp\_plant:\*\*
- 11: sp\_rodent:\*\*
- 12: sp\_virus:\*\*
- 13: sp\_vertebrate:\*\*
- 14: sp\_unclassified:\*\*
- 15: sp\_xvirus:\*\*
- 16: sp\_bacteriap:\*\*
- 17: sp\_archaeap:\*\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	48	100.0	1163	4 Q8N6U4	Q8N6U4 homo sapien
2	44	91.7	338	11 Q8BUU6	Q8BUU6 mus musculus
3	44	91.7	1222	11 Q8K173	Q8K173 mus musculus
4	44	91.7	1464	11 Q8BKV2	Q8BKV2 mus musculus
5	44	91.7	1464	11 Q8BLW4	Q8BLW4 mus musculus
6	44	91.7	1464	11 Q7TT32	Q7TT32 mus musculus
7	41	85.4	379	16 Q92Q09	Q92Q09 rhizobium m
8	39	81.2	372	16 Q87101	Q87101 vibrio para
9	38	79.2	210	17 Q8PXL0	Q8PXL0 methanoharc
10	38	79.2	383	17 Q9YFP0	Q9YFP0 aeropyrum p
11	38	79.2	415	10 Q8SIV9	Q8SIV9 oryza sativ
12	38	79.2	460	2 Q7WV41	Q7WV41 pseudomonas
13	38	79.2	504	5 Q8WS58	Q8WS58 branchiosto
14	38	79.2	525	5 Q8WS58	Q8WS58 branchiosto
15	38	79.2	696	10 Q7XPP9	Q7XPP9 oryza sativ
16	37	77.1	378	2 Q93IK5	Q93IK5 vibrio sp.

17	37	77.1	476	16 Q8Z1I8	Q8Z1I8 salmonella
18	36	75.0	80	10 Q8H778	Q8H778 arabidopsis
19	36	75.0	105	10 Q8H797	Q8H797 arabidopsis
20	36	75.0	167	5 O15729	O15729 entamoeba h
21	36	75.0	322	2 O06441	O06441 rhodococcus
22	36	75.0	338	16 Q8ZR45	Q8ZR45 streptomyce
23	36	75.0	379	16 Q8UFD4	Q8UFD4 agrobacteri
24	36	75.0	381	16 Q8D7G3	Q8D7G3 vibrio vuln
25	36	75.0	392	17 Q9HIK2	Q9HIK2 thermoplas
26	36	75.0	397	16 Q831N1	Q831N1 shigella fl
27	36	75.0	404	16 Q8X554	Q8X554 escherichia
28	36	75.0	404	16 Q8CVG5	Q8CVG5 escherichia
29	36	75.0	404	16 Q7UQA6	Q7UQA6 shigella fl
30	36	75.0	452	10 Q9LW52	Q9LW52 arabidopsis
31	36	75.0	457	16 Q889B3	Q889B3 pseudomonas
32	36	75.0	506	2 Q929I8	Q929I8 oleomonas s
33	36	75.0	547	16 Q7UMV0	Q7UMV0 rhodopirell
34	36	75.0	629	6 Q29426	Q29426 oryctolagus
35	36	75.0	653	6 Q95L99	Q95L99 canis famil
36	36	75.0	1115	12 Q7TGB7	Q7TGB7 canine coro
37	36	75.0	1480	12 Q7TE92	Q7TE92 canine coro
38	36	75.0	1481	12 Q7TE91	Q7TE91 canine coro
39	35	72.9	101	10 Q9LGG6	Q9LGG6 oryza sativ
40	35	72.9	194	17 Q970Q3	Q970Q3 sulfolobus
41	35	72.9	222	5 Q9GZ20	Q9GZ20 plasmodium
42	35	72.9	250	10 Q84JP6	Q84JP6 arabidopsis
43	35	72.9	284	10 Q94CI8	Q94CI8 lycopersico
44	35	72.9	292	16 Q8D7T5	Q8D7T5 vibrio vuln
45	35	72.9	302	16 Q8FPY3	Q8FPY3 corynebacte

#### ALIGNMENTS

RESULT 1  
Q8N6U4  
ID Q8N6U4  
AC Q8N6U4; PRELIMINARY; PRT; 1163 AA.  
DT 01-OCT-2002 (Tremblrel. 22, Created)  
DT 01-OCT-2002 (Tremblrel. 22, Last sequence update)  
DE 01-OCT-2003 (Tremblrel. 25, Last annotation update)  
DE Similar to collagen, type III, alpha 1 (Ehlers-Danlos syndrome type IV, autosomal dominant).  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Liver;  
RA Strausberg R.;  
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; BC028178; AAH28178.1; -  
DR GO; GO:0005581; C:collagen; IEA.  
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.  
DR InterPro; IPR008160; Collagen.  
DR InterPro; IPR002181; Fibrinogen C.  
DR InterPro; IPR001007; VWF\_C.  
DR Pfam; PF01410; COLFI; 1.  
DR Pfam; PF01391; Collagen; 13.  
DR ProDom; PD002078; Fib\_collagen\_C; 1.  
DR SMART; SM00038; COLFI; 1.  
DR SMART; SM00214; WVC; 1.  
DR PROSITE; PS01208; WVC\_1; 1.  
DR PROSITE; PSS0184; WVC\_2; 1.  
KW Collagen.  
SQ SEQUENCE 1163 AA; 111899 MW; 9E0C6BE1E94D6357 CRC64;

Query Match 100.0%; Score 48; DB 4; Length 1163;  
Best Local Similarity 100.0%; Pred. No. 8.2;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 IGGEKAGGF 9
Db 905 IGGEKAGGF 913

RESULT 2
Q8BJU6 PRELIMINARY; PRT; 338 AA.
AC Q8BJU6;
DT 01-MAR-2003 (T-EMBLrel. 23, Created)
DT 01-MAR-2003 (T-EMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE Collagen alpha 1 (Fragment).
GN COL3A1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Embryo;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium
RA "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs."
RL Nature 420:563-573(2002).
DR EMBL; AK079113; BAC37545.1; -.
DR MGD; MGI:88453; Col3a1.
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR002181; Fibrinogen_C.
DR Pfam; PF01410; COLFI; 1.
DR Pfam; PF01391; Collagen; 18.
DR ProDom; PD002078; Fib_collagen_C; 1.
DR SMART; SM00038; COLFI; 1.
FT NON TER 1
SQ SEQUENCE 338 AA; 35813 MW; 15BE369D8690F37A CRC64;

Query Match 91.7%; Score 44; DB 11; Length 338;
Best Local Similarity 77.8%; Pred. No. 11;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 IGGEKAGGF 9
Db 80 VGGEKSGGF 88

RESULT 3
Q8K173 PRELIMINARY; PRT; 1222 AA.
AC Q8K173;
DT 01-OCT-2002 (T-EMBLrel. 22, Created)
DT 01-OCT-2002 (T-EMBLrel. 22, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE Hypothetical protein (Fragment).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Breast tumor;
RA Strausberg R.;
RA Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC028248; AAH28248.1; -.
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR InterPro; IPR008161; Clg helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR002181; Fibrinogen_C.

QY 1 IGGEKAGGF 9
Db 1206 VGGEKSGGF 1214

RESULT 4
Q8BKX2 PRELIMINARY; PRT; 1464 AA.
AC Q8BKX2;
DT 01-MAR-2003 (T-EMBLrel. 23, Created)
DT 01-MAR-2003 (T-EMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE Collagen alpha 1.
GN COL3A1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Head;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium
RA "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs."
RL Nature 420:563-573(2002).
DR EMBL; AK048546; BAC33370.1; -.
DR MGD; MGI:88453; Col3a1.
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR InterPro; IPR008161; Clg helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR002181; Fibrinogen_C.
DR InterPro; IPR000885; Fib_collagen_C.
DR InterPro; IPR001007; VWF_C.
DR Pfam; PF01410; COLFI; 1.
DR Pfam; PF01391; Collagen; 18.
DR ProDom; PD000007; Clg_helix; 1.
DR ProDom; PD002078; Fib_collagen_C; 1.
DR SMART; SM00038; COLFI; 1.
DR SMART; SM00214; VWC; 1.
DR PROSITE; PS01208; VWFC.1; 1.
DR PROSITE; PS50184; VWFC.2; 1.
SQ SEQUENCE 1464 AA; 138947 MW; 1E4ED95399EF42C12 CRC64;

Query Match 91.7%; Score 44; DB 11; Length 1464;
Best Local Similarity 77.8%; Pred. No. 53;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 IGGEKAGGF 9
Db 1206 VGGEKSGGF 1214

RESULT 5
Q8BLW4 PRELIMINARY; PRT; 1464 AA.
AC Q8BLW4;

```

```
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Collagen alpha 1.
GN COL3A1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Aorta, and Vein;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
DR EMBL; AK041115; BAC30826.1; -.
DR MGD; MGI:88453; Col3a1.
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR InterPro; IPR008161; C1g_helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR002181; Fibrinogen C.
DR InterPro; IPR000885; Fib collagen_C.
DR InterPro; IPR001007; VWF_C.
DR Pfam; PF01410; COLFI; 1-.
DR Pfam; PF01391; Collagen; 18.
DR ProDom; PD000007; C1g_helix; 1.
DR ProDom; PD002078; Fib collagen_C; 1.
DR SMART; SM00038; COLFI; 1.
DR SMART; SM00214; VWC; 1.
DR PROSITE; PS01208; VWF_C; 1.
DR PROSITE; PS0184; VWF_C; 1.
SQ SEQUENCE 1464 AA; 138939 MW; 91F3246D90818449 CRC64;

Query Match 91.7%; Score 44; DB 11; Length 1464;
Best Local Similarity 77.8%; Pred. No. 53;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 IGGEXAGGF 9
Db 1206 VGGKSGGF 1214

RESULT 6
Q7TT32
ID Q7TT32 PRELIMINARY; PRT; 1464 AA.
AC Q7TT32;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Brain;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan B., Moore T., Max S.I., Wang J., Hsieh F.,
RA Datchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Scapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udén T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
```

```
RA Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Brain;
RX Strausberg R.;
RA Submitted (May-2003) to the EMBL/GenBank/DBJ databases.
RL EMBL; BC052398; AAHS2398.1; -.
RW Hypothetical protein.
SQ SEQUENCE 1464 AA; 138971 MW; 2B38BC27AF21590B CRC64;

Query Match 91.7%; Score 44; DB 11; Length 1464;
Best Local Similarity 77.8%; Pred. No. 53;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 IGGEXAGGF 9
Db 1206 VGGKSGGF 1214

RESULT 7
Q92Q09
ID Q92Q09 PRELIMINARY; PRT; 379 AA.
AC Q92Q09;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Probable aminomethyltransferase (Glycine cleavage system T protein)
DE (EC 2.1.2.10).
GN GCVT OR R01549 OR SMC02047.
OS Rhizobium meliloti (Sinorhizobium meliloti).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Sinorhizobium/Ensifer group; Sinorhizobium.
OX NCBI_TaxID=382;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1021;
RX MEDLINE=21396507; PubMed=11481430;
RA Capela D., Barloy-Hubler F., Gouzy J., Bothe G., Ampe F., Batut J.,
RA Boistard P., Becker A., Boutry M., Cadieu E., Dreano S., Gloux S.,
RA Godrie T., Goffeau A., Kahn D., Kiss E., Lelaure V., Masuy D.,
RA Pohl T., Portetelle D., Puehler A., Purnelle B., Ramsperger U.,
RA Renard C., Thebault P., Vandenbol M., Weidner S., Galibert F.;
RT "Analysis of the chromosome sequence of the legume symbiont
RT Sinorhizobium meliloti strain 1021.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882(2001).
RN EMBL; AL591787; CAC46128.1; -.
DR GO; GO:0004047; F:aminomethyltransferase activity; IEA.
DR GO; GO:0004374; F:glycine cleavage system; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR GO; GO:0006546; P:glycine catabolism; IEA.
DR InterPro; IPR006223; Gcvt.
DR InterPro; IPR006222; GCV_T.
DR Pfam; PF01571; GCV_T; 1.
DR TIGRFAMs; TIGR00528; gcvT; 1.
RW Transferase; Complete proteome.
SQ SEQUENCE 379 AA; 40369 MW; 047C7002EFA30CF CRC64;

Query Match 85.4%; Score 41; DB 16; Length 379;
Best Local Similarity 87.5%; Pred. No. 43;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 GGBKAGGF 9
Db 270 GGBKAGGF 277
```

Mon Sep 20 11:05:45 2004

us-10-615-959-41.rspt

```

RESULT 8
Q87101 PRELIMINARY; PRT; 372 AA.
ID Q87101
AC Q87101
DT 01-JUN-2003 (TReMBLrel. 24, Created)
DT 01-JUN-2003 (TReMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE Glycine cleavage system protein T2.
DE VPA0805.
GN Vibrio parahaemolyticus.
OS Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC Vibrionaceae; Vibrio.
OX NCBI_TaxID=670;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RMD 2210633 / Serotype O3:K6;
RX MEDLINE=22508454; PubMed=12620739;
RA Makino K., Oshima K., Kurokawa K., Yokoyama K., Uda T., Tagomori K.,
RA Iijima Y., Najima M., Nakano M., Yamashita A., Kubota Y., Kimura S.,
RA Yasunaga T., Honda T., Shinagawa H., Hattori M., Iida T.;
RT "Genome sequence of Vibrio parahaemolyticus: a pathogenic mechanism
RT distinct from that of V. cholerae.";
RL Lancet 361:743-749 (2003).
DR EMBL; AF005086; BAC62148.1; -.
DR GO; GO:0004047; F:aminomethyltransferase activity; IEA.
DR InterPro; IPR006222; GCV_T.
DR Pfam; PF01571; GCV_T; 1.
KW Complete proteome.
SQ SEQUENCE 372 AA; 40322 MW; 51FC31A1D424D57 CRC64;

Query Match 81.2%; Score 39; DB 16; Length 372;
Best Local Similarity 66.7%; Pred. No. 96;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 IGGEKAGGF 9
Db 262 VGGEREGGF 270

RESULT 9
ID Q8PXLO PRELIMINARY; PRT; 210 AA.
AC Q8PXLO
DT 01-OCT-2002 (TReMBLrel. 22, Created)
DT 01-OCT-2002 (TReMBLrel. 22, Last sequence update)
DT 01-OCT-2002 (TReMBLrel. 22, Last annotation update)
DE Conserved protein.
GN MM1208.
OS Methanosarcina mazei (Methanosarcina frisia).
OC Archaea; Euryarchaeota; Euryarchaeota orders incertae sedis;
OC Methanosarcinales; Methanosarcinaceae; Methanosarcina.
OX NCBI_TaxID=2209;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Goe1 / Goe1 / ATCC BAA-199 / DSM 3647 / OCM 88;
RX MEDLINE=22120827; PubMed=12125824;
RA Deppeleier U., Johann A., Hartsch T., Merkl R., Schmitz R.A.,
RA Martinez-Arias R., Henne A., Wiersema M., Baeumer S., Jacobi C.,
RA Brueggemann H., Lienard T., Christmann A., Boemcke M., Steckel S.,
RA Bhattacharya A., Lykidis A., Overbeek R., Klenk H.-P., Gunsalus R.P.,
RA Fritz H.-J., Gottschalk G.;
RT "The genome of Methanosarcina mazei: evidence for lateral gene
RT transfer between Bacteria and Archaea.";
RL J. Mol. Microbiol. Biotechnol. 4:453-461 (2002).
DR EMBL; AE013351; AM30904.1; -.
KW Complete proteome.
SQ SEQUENCE 210 AA; 24264 MW; F53D2DLDB598AAFO CRC64;

Query Match 79.2%; Score 38; DB 17; Length 210;
Best Local Similarity 100.0%; Pred. No. 79;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 GGEKAGGF 9
Db 39 GGEKAGGF 45

RESULT 10
Q9YFPO PRELIMINARY; PRT; 383 AA.
ID Q9YFPO
AC Q9YFPO
DT 01-NOV-1999 (TReMBLrel. 12, Created)
DT 01-NOV-1999 (TReMBLrel. 12, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE 383AA long hypothetical S2P metalloprotease.
DE APE0209.
GN Aeropyrum pernix.
OS Archaea; Crenarchaeota; Thermoprotei; Desulfurococcales;
OC Desulfurococaceae; Aeropyrum.
OX NCBI_TaxID=56636;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=X1;
RX MEDLINE=99110339; PubMed=10382966;
RA Kawarabayashi Y., Hino Y., Horikawa H., Yamazaki S., Haikawa Y.,
RA Jin-no K., Takahashi M., Sekine M., Baba S.-I., Ankai A., Kosugi H.,
RA Hosoyama A., Fukui S., Nagai Y., Nishijima K., Nakazawa H.,
RA Takamiya M., Masuda S., Funahashi T., Tanaka T., Kudoh Y.,
RA Yamazaki J., Kishida N., Ogunishi A., Aoki K.-I., Kubota K.,
RA Nakamura Y., Nomura N., Sako Y., Kikuchi H.;
RT "Complete genome sequence of an aerobic hyper-thermophilic
RT crenarchaeon, Aeropyrum pernix X1.";
RL DNA Res. 6:83-101 (1999).
DR EMBL; AP000058; BAA79121.1; -.
DR PIR; G72777; G72777.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0004222; F:metalloendopeptidase activity; IEA.
DR GO; GO:0008270; F:zinc ion binding; IEA.
DR GO; GO:0007242; P:intracellular signaling cascade; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR008915; Peptidase M50.
DR InterPro; IPR001193; Pept_M50_SREBP.
DR InterPro; IPR006025; Pept_M_Zn_BS.
DR Pfam; PF00595; PDZ; 1.
DR Pfam; PF02163; Peptidase M50; 1.
DR PRINTS; PR01000; SREBPS2PTASE.
DR SMART; SM00228; PDZ; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
KW Protease; Metalloprotease; Complete proteome.
SQ SEQUENCE 383 AA; 39729 MW; FE7289C4F068F76 CRC64;

Query Match 79.2%; Score 38; DB 17; Length 383;
Best Local Similarity 87.5%; Pred. No. 1.5e+02;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 GGEKAGGF 9
Db 99 GGEKAGGF 106

RESULT 11
Q8SIV9 PRELIMINARY; PRT; 415 AA.
ID Q8SIV9
AC Q8SIV9
DT 01-JUN-2002 (TReMBLrel. 21, Created)
DT 01-JUN-2002 (TReMBLrel. 21, Last sequence update)
DT 01-JUN-2003 (TReMBLrel. 24, Last annotation update)
DE Putative beta-amylase.
DE P0503C12.23.
GN Oryza sativa (japonica cultivar-group).
OS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzeae; Oryza.
OX NCBI_TaxID=39947;

```



```

RN RP SEQUENCE FROM N.A.
RC STRAIN=cv. Nipponbare;
RA "Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
RT clone:PO503C12.";
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP003268; BAB89698.1; -.
DR Gramene; Q8S1V9; -.
DR GO; GO:0016161; F:beta-amylase activity; IEA.
DR GO; GO:0004289; F:subtilase activity; IEA.
DR GO; GO:0000272; P:polysaccharide catabolism; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR001554; Glyco_hydro_14.
DR InterPro; IPR000209; Peptidase_S8.
DR Pfam; PF01373; Glyco_hydro_14; 2.
DR PROSITE; PS00136; SUBTILASE_ASP; 1.
SQ SEQUENCE 415 AA; 46150 MW; 851DEBA6BDF2E7BA CRC64;

Query Match 79.2%; Score 38; DB 10; Length 415;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GGEKAGG 8
DB 141 GGEKAGG 147

RESULT 12
Q7WY41 PRELIMINARY; PRT; 460 AA.
AC Q7WY41;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
GN RL016.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PA14;
RA He J., Rahme L.G.;
RT "Pseudomonas aeruginosa PA14 pathogenicity island (PAPI) 1.";
RL Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY273869; AAP84143.1; -.
DR Hypothetical protein.
SQ SEQUENCE 460 AA; 48289 MW; A56F37B515F2F28D CRC64;

Query Match 79.2%; Score 38; DB 2; Length 460;
Best Local Similarity 75.0%; Pred. No. 1.8e+02;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 IGGEKAGG 8
DB 199 VGDKAGG 206

RESULT 13
Q8WS58 PRELIMINARY; PRT; 504 AA.
AC Q8WS58;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Presenilin.
GN PS.
OS Branchiostoma floridae (Florida lancelet) (Amphioxus).
OC Eukaryota; Metazoa; Chordata; Cephalochordata; Branchiostomidae;
OC Branchiostoma.
OX NCBI_TaxID=7739;

```

```

RN RP SEQUENCE FROM N.A.
RC MEDLINE=21590369; PubMed=11733140;
RA Martinez-Mir A., Canestro C., Gonzalez-Duarte R., Albalat R.;
RT "Characterization of the amphioxus presenilin gene in a high gene-
RT density genomic region illustrates duplication during the vertebrate
RT lineage.";
RL Gene 279:157-164 (2001).
DR EMBL; AF369891; AAL40416.1; -.
DR EMBL; AF369890; AAL40416.1; JOINED.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0003700; F:transcription factor activity; IEA.
DR GO; GO:0007242; P:intracellular signaling cascade; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR000847; HTH_LYSR.
DR InterPro; IPR006639; Peptidase_A22.
DR Pfam; PF01080; Presenilin; 1.
DR PRINTS; PR01072; PRESENILIN.
DR SMART; SM00730; PSN; 1.
DR PROSITE; PS00044; HTH_LYSR_FAMILY; 1.
SQ SEQUENCE 504 AA; 55443 MW; 9C3794D0302859B4 CRC64;

Query Match 79.2%; Score 38; DB 5; Length 504;
Best Local Similarity 66.7%; Pred. No. 2e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 IGGEKAGG 9
DB 343 VGEEGGF 351

RESULT 14
Q8WS59 PRELIMINARY; PRT; 525 AA.
AC Q8WS59;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Presenilin.
GN PS.
OS Branchiostoma floridae (Florida lancelet) (Amphioxus).
OC Eukaryota; Metazoa; Chordata; Cephalochordata; Branchiostomidae;
OC Branchiostoma.
OX NCBI_TaxID=7739;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=21590369; PubMed=11733140;
RA Martinez-Mir A., Canestro C., Gonzalez-Duarte R., Albalat R.;
RT "Characterization of the amphioxus presenilin gene in a high gene-
RT density genomic region illustrates duplication during the vertebrate
RT lineage.";
RL Gene 279:157-164 (2001).
DR EMBL; AF369891; AAL40416.1; -.
DR EMBL; AF369890; AAL40416.1; JOINED.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0003700; F:transcription factor activity; IEA.
DR GO; GO:0007242; P:intracellular signaling cascade; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR000847; HTH_LYSR.
DR InterPro; IPR006639; Peptidase_A22.
DR Pfam; PF01080; Presenilin; 1.
DR PRINTS; PR01072; PRESENILIN.
DR SMART; SM00730; PSN; 1.
DR PROSITE; PS00044; HTH_LYSR_FAMILY; 1.
SQ SEQUENCE 525 AA; 57598 MW; 2B14CF77A80F07DE CRC64;

Query Match 79.2%; Score 38; DB 5; Length 525;
Best Local Similarity 66.7%; Pred. No. 2.1e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 IGGEKAGG 9

```

Db 364 VGEDEGGF 372  
:||||: |||

RESULT 15

Q7XPP9 PRELIMINARY; PRT; 696 AA.  
AC Q7XPP9;  
DT 01-OCT-2003 (TREMELrel. 25, Created)  
DT 01-OCT-2003 (TREMELrel. 25, Last sequence update)  
DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)  
DE OSJNBa0053K19.24 protein.  
DE OSJNBa0053K19.24.  
GN OSJNBa0053K19.24.  
OS Oryza sativa (Rice).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
OC Ehrhartoideae; Oryzeae; Oryza.  
OX NCBI\_TaxID=4530;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Han B., Feng Q., Huang Y.C., Li Y., Zhu J.J., Zhao Q., Hu X.,  
RA Liu Y.L., Mu J., Yu Z., Chen L., Fan D.L., Weng Q.J., Zhang L.,  
RA Lu Y.Q., Yu S.L., Liu X.H., Lu T.T., Zhang Y.J., Lu Y., Li C., Li T.,  
RA Zhang Y., Hu H., Jia P.X., Qian Y.M., Ying K., Zhou B., Chen Z.H.,  
RA Hao P., Zhang L., Wu M., Zhang R.Q., Guan J.P., Fu G., Wang S.Y.,  
RA Ren S.X., Lv G., Lin W., Gu W.Q., Zhu G.F., Tu Y.F., Jia J., Yin H.F.,  
RA Zhang Y., Cai Z., Chen J., Kang H., Chen X.Y., Shao C.Y., Sun Y.,  
RA Hu Q.P., Zhang X.L., Zhang W., Wang L.J., Ding C.W., Sheng H.H.,  
RA Gu J.L., Chen S.T., Ni L., Zhu F.H., Hong G.F.;  
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AL606645; CA603516.1; -.  
SQ SEQUENCE 696 AA; 77629 MW; 5E506EC0221AF5EC CRC64;

Query Match 79.2%; Score 38; DB 10; Length 696;  
Best Local Similarity 100.0%; Pred. No. 2.8e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 GGEKAGG 8  
|||||  
Db 72 GGEKAGG 78

Search completed: September 18, 2004, 04:29:39  
Job time : 33.9412 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 18, 2004, 03:47:39 ; Search time 58.8824 Seconds  
(without alignments)  
52.784 Million cell updates/sec

Title: US-10-615-959-42

Perfect score: 56  
Sequence: 1 IAGIGEXAGG 11

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_29Jan04.\*  
1: Geneseq1980s.\*  
2: Geneseq1990s.\*  
3: Geneseq2000s.\*  
4: Geneseq2001s.\*  
5: Geneseq2002s.\*  
6: Geneseq2003as.\*  
7: Geneseq2003bs.\*  
8: Geneseq2004s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	56	100.0	11	4	AAB61742 Human typ
2	56	100.0	12	4	AAB61743 Human typ
3	56	100.0	25	7	Aae38632 Human C-T
4	56	100.0	262	7	Aae38640 Human col
5	56	100.0	270	2	AAY07375 Procollag
6	56	100.0	293	6	Abu70813 Human adi
7	56	100.0	309	4	Aau23675 Novel hum
8	56	100.0	309	4	Abg60248 Human ova
9	56	100.0	339	6	Abg61719 Novel ova
10	56	100.0	623	2	AAW12843 Human adi
11	56	100.0	842	7	Ade08475 Novel pro
12	56	100.0	1466	5	Abb50291 Collagen
13	56	100.0	1466	6	Abu54454 Human tum
14	56	100.0	1466	6	Abu54454 Human tum
15	56	100.0	1466	6	Abu54454 Human tum
16	56	100.0	1466	6	Abu54454 Human tum
17	56	100.0	1466	6	Abu54454 Human tum
18	56	100.0	1466	6	Abu54454 Human tum
19	56	100.0	1466	6	Abu54454 Human tum
20	56	100.0	1466	6	Abu54454 Human tum
21	55	98.2	1466	4	Aae02537 Porcine a
22	50	89.3	1078	2	AAW17104 Collagen
23	50	89.3	1078	3	AAY96125 Collagen
24	50	89.3	1078	5	AAE16478 Collagen
25	50	89.3	1078	5	ABb80736 Collagen

26	50	89.3	1078	5	ABB09628 Amino aci
27	49	87.5	1466	4	AAE02534 Bovine al
28	49	87.5	1466	4	AAE02533 Bovine al
29	48	85.7	30	2	AAE02068 Recombina
30	42	75.0	8	4	AAE02068 Recombina
31	42	75.0	9	4	AAE02068 Recombina
32	42	75.0	11	3	AAE02068 Recombina
33	42	75.0	11	3	AAE02068 Recombina
34	42	75.0	11	3	AAE02068 Recombina
35	42	75.0	11	3	AAE02068 Recombina
36	42	75.0	11	3	AAE02068 Recombina
37	39	69.6	240	3	AAE02068 Recombina
38	39	69.6	247	2	AAE02068 Recombina
39	39	69.6	253	6	AAE02068 Recombina
40	39	69.6	304	2	AAE02068 Recombina
41	39	69.6	575	5	AAE02068 Recombina
42	39	69.6	674	2	AAE02068 Recombina
43	39	69.6	846	6	AAE02068 Recombina
44	39	69.6	846	7	AAE02068 Recombina
45	39	69.6	1114	4	AAE02068 Recombina

## ALIGNMENTS

RESULT 1  
AAB61742  
ID AAB61742 standard; peptide; 11 AA.  
XX  
AC AAB61742;  
XX  
DT 20-APR-2001 (first entry)  
XX  
DE Human type III collagen carboxy-telopeptide fragment.  
XX  
KW Type II collagen; immunoassay; cartilage; telopeptide; human;  
KW Type III collagen.  
XX  
OS Synthetic.  
OS Homo sapiens.  
XX  
PN WO200079284-A1.  
XX  
PD 28-DEC-2000.  
XX  
PF 10-DEC-1999; 99WO-US029357.  
XX  
PR 17-JUN-1999; 99US-00335098.  
PR 29-JUN-1999; 99US-0141574P.  
PR 02-JUL-1999; 99US-0142274P.  
PR 07-JUL-1999; 99US-0142675P.  
PR 30-AUG-1999; 99US-00385740.  
XX  
(WASH-) WASHINGTON RES FOUND.  
XX  
PI Eyre DR;  
XX  
WPI; 2001-146859/15.  
XX  
Assay for detecting cross-linked telopeptide analytes indicative of type II collagen resorption in vivo in a body fluid sample, comprises contacting the sample with an antibody which binds to the analyte.  
XX  
Disclosure; Page 15; 34pp; English.  
XX  
The invention relates to immunoassays for measuring type II collagen (cartilage) resorption in vivo. The method of analysing a body fluid sample for the presence of an analyte indicative of a physiological condition, involves contacting the body fluid sample with an antibody (Ab) which binds to the analyte, detecting binding of Ab in the body fluid sample, and correlating any detected binding to the physiological condition. The analysis is useful for measuring type II collagen (cartilage) resorption in vivo, for distinguishing between resorption of

Mon Sep 20 11:05:46 2004

CC non-mineralized and mineralized cartilage, and for measuring total  
 CC cartilage resorption in vivo. The present sequence represents a linear  
 CC synthetic peptide based on the carboxy-terminal telopeptide sequence of  
 CC human collagen type III  
 XX  
 SQ Sequence 11 AA;  
 Query Match 100.0%; Score 56; DB 4; Length 11;  
 Best Local Similarity 100.0%; Pred. No. 0.016; 0; Gaps 0;  
 Matches 11; Conservative 0; Mismatches 0; Indels 0;  
 QY 1 IAGIGGEKAGG 11  
 DB |||||  
 1 IAGIGGEKAGG 11  
 RESULT 2  
 AAB61743  
 ID AAB61743 standard; peptide; 12 AA.  
 XX  
 AC AAB61743;  
 XX  
 DT 20-APR-2001 (first entry)  
 XX  
 DE Human type III collagen carboxy-telopeptide fragment.  
 XX  
 KW Type II collagen; immunoassay; cartilage; telopeptide; human;  
 KW type III collagen.  
 XX  
 OS Synthetic.  
 OS Homo sapiens.  
 XX  
 PN WO200079284-A1.  
 XX  
 EN 28-DEC-2000.  
 XX  
 PD 10-DEC-1999; 99WO-US029357.  
 XX  
 PF 17-JUN-1999; 99US-00335098.  
 PR 29-JUN-1999; 99US-0141574P.  
 PR 02-JUL-1999; 99US-0142274P.  
 PR 07-JUL-1999; 99US-0142675P.  
 PR 30-AUG-1999; 99US-00385740.  
 XX  
 PA (WASH-) WASHINGTON RES FOUND.  
 XX  
 PT Eyre DR;  
 XX  
 DR WPT; 2001-146859/15.  
 XX  
 PT Assay for detecting cross-linked telopeptide analytes indicative of type  
 PT II collagen resorption in vivo in a body fluid sample, comprises  
 PT contacting the sample with an antibody which binds to the analyte.  
 XX  
 PS Disclosure; Page 15; 34pp; English.  
 XX  
 CC The invention relates to immunoassays for measuring type II collagen  
 CC (cartilage) resorption in vivo. The method of analysing a body fluid  
 CC sample for the presence of an analyte indicative of a physiological  
 CC condition, involves contacting the body fluid sample with an antibody  
 CC (Ab) which binds to the analyte, detecting binding of Ab in the body  
 CC fluid sample, and correlating any detected binding to the physiological  
 CC condition. The analysis is useful for measuring type II collagen  
 CC (cartilage) resorption in vivo, for distinguishing between resorption of  
 CC non-mineralized and mineralized cartilage, and for measuring total  
 CC cartilage resorption in vivo. The present sequence represents a linear  
 CC synthetic peptide based on the carboxy-terminal telopeptide sequence of  
 CC human collagen type III  
 XX  
 SQ Sequence 12 AA;  
 Query Match 100.0%; Score 56; DB 4; Length 12;  
 Best Local Similarity 100.0%; Pred. No. 0.016;

Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 IAGIGGEKAGG 11  
 DB |||||  
 1 IAGIGGEKAGG 11  
 RESULT 3  
 AAE38632  
 ID AAE38632 standard; peptide; 25 AA.  
 XX  
 AC AAE38632;  
 XX  
 DT 04-DEC-2003 (first entry)  
 XX  
 DE Human C-terminal telopeptide alpha 1 #4.  
 XX  
 KW Human; therapy; osteoporosis; left-ventricular hypertrophy; scleroderma;  
 KW liver cirrhosis; congestive heart failure; multiple sclerosis; fibrosis;  
 KW coal-miner's pneumoconiosis; muscle-wasting syndrome; diabetes mellitus;  
 KW rheumatoid arthritis; Alzheimer's disease; muscular dystrophy; cancer;  
 KW telopeptide alpha.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO2003068919-A2.  
 XX  
 PD 21-AUG-2003.  
 XX  
 PF 12-FEB-2003; 2003WO-US004183.  
 PR 12-FEB-2002; 2002US-0356008P.  
 XX  
 PA (REGC ) UNIV CALIFORNIA.  
 XX  
 PI Hellerstein MK;  
 XX  
 DR WPI; 2003-689661/65.  
 XX  
 PT Determining rate of biosynthesis or breakdown of inaccessible biological  
 PT molecules, useful e.g. for diagnosis or monitoring treatment, by  
 PT administering labeled precursor.  
 XX  
 PS Claim 24; Page 99; 105pp; English.  
 XX  
 CC The invention relates to a method of determining the rate of biosynthesis  
 CC or breakdown of at least one inaccessible biological molecule in a  
 CC subject. The method is useful for diagnosis or monitoring and treatment  
 CC of diseases associated with an altered rate of biosynthesis/breakdown of  
 CC an isotopically labelled precursor molecule, specifically osteoporosis;  
 CC left-ventricular hypertrophy; liver cirrhosis or fibrosis; congestive  
 CC heart failure; scleroderma; coal-miner's pneumoconiosis; cardiac or lung  
 CC fibrosis; Alzheimer's disease; multiple sclerosis; rheumatoid arthritis;  
 CC diabetes mellitus; muscle-wasting syndromes; muscular dystrophy; athletic  
 CC training and cancer. The method is also useful for screening candidate  
 CC gene or protein targets, phenotypic/human validation studies on potential  
 CC drugs, drug mechanism studies and determining the risk of developing the  
 CC disease. The present sequence is human C-terminal telopeptide alpha 1.  
 CC This sequence is used to illustrate the method of the invention  
 XX  
 SQ Sequence 25 AA;  
 Query Match 100.0%; Score 56; DB 7; Length 25;  
 Best Local Similarity 100.0%; Pred. No. 0.038;  
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 IAGIGGEKAGG 11  
 DB |||||  
 9 IAGIGGEKAGG 19  
 RESULT 4  
 AAE38640

ID AAE38640 standard; protein; 262 AA.  
 AC AAE38640;  
 XX  
 DT 04-DEC-2003 (first entry)  
 XX  
 DE Human collagen type I cross-linked C-terminal peptide, PIIICP(alphal).  
 XX  
 KW Human; therapy; osteoporosis; left-ventricular hypertrophy; scleroderma;  
 KW liver cirrhosis; congestive heart failure; multiple sclerosis; scleroderma;  
 KW coal-miner's pneumoconiosis; muscle-wasting syndrome; diabetes mellitus;  
 KW rheumatoid arthritis; Alzheimer's disease; muscular dystrophy; cancer;  
 KW collagen type I cross-linked carboxy-terminal peptide.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO2003068919-A2.  
 XX  
 PD 21-AUG-2003.  
 XX  
 PF 12-FEB-2003; 2003WO-US004183.  
 XX  
 PR 12-FEB-2002; 2002US-0356008P.  
 XX  
 PA (REGC ) UNIV CALIFORNIA.  
 XX  
 PI Hellerstein MK;  
 XX  
 DR WPI; 2003-689661/65.  
 XX  
 PT Determining rate of biosynthesis or breakdown of inaccessible biological  
 PT molecules, useful e.g. for diagnosis or monitoring treatment, by  
 PT administering labeled precursor.  
 XX  
 PS Claim 24; Page 102-103; 105pp; English.  
 XX  
 CC The invention relates to a method of determining the rate of biosynthesis  
 CC or breakdown of at least one inaccessible biological molecule in a  
 CC subject. The method is useful for diagnosis or monitoring and treatment  
 CC of diseases associated with an altered rate of biosynthesis/breakdown of  
 CC an isotopically labelled precursor molecule, specifically osteoporosis;  
 CC left-ventricular hypertrophy; liver cirrhosis or fibrosis; congestive  
 CC heart failure; scleroderma; coal-miner's pneumoconiosis; cardiac or lung  
 CC fibrosis; Alzheimer's disease; multiple sclerosis; rheumatoid arthritis;  
 CC diabetes mellitus; muscle-wasting syndromes; muscular dystrophy; athletic  
 CC training and cancer. The method is also useful for screening candidate  
 CC gene or protein targets, phenotypic/human validation studies on potential  
 CC drugs, drug mechanism studies and determining the risk of developing the  
 CC disease. The present sequence is human collagen type I cross-linked  
 CC carboxy terminal peptide (ICTP). This sequence is used to illustrate the  
 CC method of the invention  
 XX  
 SQ Sequence 262 AA;  
 XX  
 Query Match 100.0%; Score 56; DB 7; Length 262;  
 Best Local Similarity 100.0%; Pred. No. 0.43;  
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 IAGIGGKAGG 11  
 |||||  
 Db 1 IAGIGGKAGG 11  
 |||||  
 RESULT 5  
 AAY07375  
 ID AAY07375 standard; protein; 270 AA.  
 XX  
 AC AAY07375;  
 XX  
 DT 16-JUL-1999 (first entry)  
 XX  
 DE Procollagen-III-C-terminal propeptide.  
 XX

KW Alcoholic cirrhosis; biliary cirrhosis; hepatitis; schistosomiasis;  
 KW cardiac fibrosis; Crohn's disease; diabetic neuropathy; collagen;  
 KW fibril; procollagen-III-C-terminal propeptide; immunoassay; diagnosis;  
 KW glucocorticosteroid; antibody.  
 XX  
 OS Homo sapiens.  
 XX  
 PN EP913692-A1.  
 XX  
 PD 06-MAY-1999.  
 XX  
 PF 31-OCT-1997; 97EP-00119018.  
 XX  
 PR 31-OCT-1997; 97EP-00119018.  
 XX  
 PA (FARB ) BAYER AG.  
 XX  
 PI Burchardt ER, Kroll W, Neumann R, Schroeder W;  
 XX  
 DR WPI; 1999-256748/22.  
 DR N-PSDB; AAX57501.  
 XX  
 PT New immunoassay for procollagen-III-C-terminal propeptide (PIICP) useful  
 PT for diagnosing diseases such as alcoholic cirrhosis, hepatitis and  
 PT Duchenne's muscular dystrophy.  
 XX  
 PS Disclosure; Fig 1; 28pp; English.  
 XX  
 CC Diseases such as alcoholic cirrhosis, biliary cirrhosis, hepatitis,  
 CC schistosomiasis, cardiac fibrosis, Crohn's disease, diabetic neuropathy  
 CC and fibroses caused by surgery result from an inappropriate production of  
 CC collagen fibrils. Collagen III is synthesised as a preproprotein which is  
 CC modified by post-translation modification. The mature collagen III  
 CC fibrils are generated by cleavage of a propeptide from the C-terminus  
 CC followed by cleavage of a fragment from the N-terminus. This sequence  
 CC corresponds to the procollagen III C-terminal propeptide (PIICP). As the  
 CC fibrils are formed, the PIICP propeptide is released into the  
 CC surrounding tissue fluids. The invention relates to an immunoassay to  
 CC determine PIICP concentrations and thus diagnose the above diseases and  
 CC others which result from inappropriate collagen deposition. Different  
 CC fibrotic diseases may show different levels of PIICP in the serum so the  
 CC immunoassay may also be used to determine which disease a particular  
 CC patient is suffering from. The new immunoassay may also be used to  
 CC measure the rate of collagen synthesis in patients being treated with  
 CC glucocorticosteroids. In addition the new antibodies may be used in  
 CC immunohistochemical staining of cryostat and paraffin sections to assess  
 CC collagen synthesis in tissue samples from patients suspected of having  
 CC fibrotic disease  
 XX  
 SQ Sequence 270 AA;  
 XX  
 Query Match 100.0%; Score 56; DB 2; Length 270;  
 Best Local Similarity 100.0%; Pred. No. 0.44;  
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 IAGIGGKAGG 11  
 |||||  
 Db 9 IAGIGGKAGG 19  
 |||||  
 RESULT 6  
 ABU70813  
 ID ABU70813 standard; protein; 293 AA.  
 XX  
 AC ABU70813;  
 XX  
 DT 10-JUN-2003 (first entry)  
 XX  
 DE Human adipocyte Selected Interacting domain, SID, #444.  
 DE Human; prev; adipocyte; SID; selected interacting domain; anorectic;  
 KW antidiabetic; protein-protein interaction; diabetes;  
 KW yeast 2-hybrid assay; metabolic disorder; obesity.  
 XX

Mon Sep 20 11:05:46 2004

us-10-615-959-42.rag

inflammatory disorder; cardiovascular disorder; reproductive disorder;  
blood-related disorder; infectious disorder; cytostatic; anti arthritic;  
nephrotropic; anticoagulant.

XX Homo sapiens.  
OS WO200286122-A2.  
XX PD 31-OCT-2002.  
XX PF 14-MAR-2002; 2002WO-EP003768.  
XX PR 14-MAR-2001; 2001US-0275734P.  
XX (HYBR-) HYBRIGENICS.  
XX Legrain P, Daviet L;  
XX WPI; 2003-103412/09.  
XX N-PSDB; ACAS7357.  
XX New complex between two interacting proteins in adipocyte cells, useful  
PT for identifying selected interacting domains that modulate protein  
PT interactions, or for preventing or treating metabolic disorders such as  
PT obesity or diabetes.  
XX Claim 6; Page 258-259; 382pp; English.  
XX The invention relates to a complex between two interacting proteins in  
CC adipocyte cells, given in the specification. The proteins are identified  
CC by selecting a bait protein from a known adipocyte marker and then  
CC performing a yeast 2-hybrid selection to isolate prey proteins encoded by  
CC members of an adipocyte cDNA library. The proteins are designated SID  
CC (RTM) (selected interacting domains) proteins. Also included are a  
CC polynucleotide encoding a polypeptide in the adipocyte cells, a  
CC recombinant host cell expressing at least one of the interacting  
CC polypeptides of the complex, selecting a modulating compound in adipocyte  
CC cells, a SID (RTM) polypeptide comprising any of the 738 amino acid  
CC sequences given in the specification (including its fragment or variant),  
CC a SID (RTM) polynucleotide comprising any of the 738 nucleotide sequences  
CC given in the specification (including its fragment or variant), a vector  
CC comprising the SID (RTM) polynucleotide, a recombinant host cell  
CC comprising the vector, a protein chip comprising the polypeptides and a  
CC record comprising all or part of the data, listed in the specification.  
CC The complex, polypeptides, polynucleotides and compounds are useful for  
CC preventing or treating metabolic disorders such as obesity or diabetes.  
CC The polynucleotides are useful as probes or primers. The complex is  
CC particularly useful for identifying selected interacting domains (SID  
CC (RTM)) for screening drugs that modulate the protein interaction, thus  
CC exhibiting the therapeutic effect. The present sequence represents a SID  
CC (prey) protein of the invention  
XX SQ Sequence 293 AA;  
Query Match 100.0%; Score 56; DB 6; Length 293;  
Best Local Similarity 100.0%; Pred. No. 0.48;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 IAGIGGEKAGG 11  
DB 136 IAGIGGEKAGG 146  
RESULT 7  
AAU23675  
ID AAU23675 standard; protein; 309 AA.  
XX AC AAU23675;  
XX 18-DEC-2001 (first entry)  
XX Novel human enzyme polypeptide #761.  
XX Human; oxidoreductase enzyme; transferase; hydrolase; lyase; isomerase;  
KW ligase; hyperproliferative disorder; immunodeficiency disorder;  
KW autoimmune disorder; neurological disorder; metabolic disorder;  
KW

PR 21-SEP-2000; 2000US-0234274P.  
PR 25-SEP-2000; 2000US-0234997P.  
PR 25-SEP-2000; 2000US-0234998P.  
PR 26-SEP-2000; 2000US-0235484P.  
PR 27-SEP-2000; 2000US-0235834P.  
PR 27-SEP-2000; 2000US-0235836P.  
PR 29-SEP-2000; 2000US-0236327P.  
PR 29-SEP-2000; 2000US-0236367P.  
PR 29-SEP-2000; 2000US-0236368P.  
PR 29-SEP-2000; 2000US-0236369P.  
PR 29-SEP-2000; 2000US-0236370P.  
PR 02-OCT-2000; 2000US-0236802P.  
PR 02-OCT-2000; 2000US-0237037P.  
PR 02-OCT-2000; 2000US-0237038P.  
PR 02-OCT-2000; 2000US-0237039P.  
PR 02-OCT-2000; 2000US-0237040P.  
PR 13-OCT-2000; 2000US-0239335P.  
PR 13-OCT-2000; 2000US-0239337P.  
PR 20-OCT-2000; 2000US-0240960P.  
PR 20-OCT-2000; 2000US-0241221P.  
PR 20-OCT-2000; 2000US-0241785P.  
PR 20-OCT-2000; 2000US-0241786P.  
PR 20-OCT-2000; 2000US-0241787P.  
PR 20-OCT-2000; 2000US-0241808P.  
PR 20-OCT-2000; 2000US-0241809P.  
PR 20-OCT-2000; 2000US-0241826P.  
PR 01-NOV-2000; 2000US-0244617P.  
PR 08-NOV-2000; 2000US-0246474P.  
PR 08-NOV-2000; 2000US-0246475P.  
PR 08-NOV-2000; 2000US-0246476P.  
PR 08-NOV-2000; 2000US-0246477P.  
PR 08-NOV-2000; 2000US-0246478P.  
PR 08-NOV-2000; 2000US-0246523P.  
PR 08-NOV-2000; 2000US-0246524P.  
PR 08-NOV-2000; 2000US-0246525P.  
PR 08-NOV-2000; 2000US-0246526P.  
PR 08-NOV-2000; 2000US-0246527P.  
PR 08-NOV-2000; 2000US-0246528P.  
PR 08-NOV-2000; 2000US-0246532P.  
PR 08-NOV-2000; 2000US-0246609P.  
PR 08-NOV-2000; 2000US-0246610P.  
PR 08-NOV-2000; 2000US-0246611P.  
PR 08-NOV-2000; 2000US-0246613P.  
PR 17-NOV-2000; 2000US-0249207P.  
PR 17-NOV-2000; 2000US-0249208P.  
PR 17-NOV-2000; 2000US-0249209P.  
PR 17-NOV-2000; 2000US-0249210P.  
PR 17-NOV-2000; 2000US-0249211P.  
PR 17-NOV-2000; 2000US-0249212P.  
PR 17-NOV-2000; 2000US-0249213P.  
PR 17-NOV-2000; 2000US-0249214P.  
PR 17-NOV-2000; 2000US-0249215P.  
PR 17-NOV-2000; 2000US-0249216P.  
PR 17-NOV-2000; 2000US-0249217P.  
PR 17-NOV-2000; 2000US-0249218P.  
PR 17-NOV-2000; 2000US-0249244P.  
PR 17-NOV-2000; 2000US-0249245P.  
PR 17-NOV-2000; 2000US-0249264P.  
PR 17-NOV-2000; 2000US-0249265P.  
PR 17-NOV-2000; 2000US-0249266P.  
PR 17-NOV-2000; 2000US-0249297P.  
PR 17-NOV-2000; 2000US-0249300P.  
PR 01-DEC-2000; 2000US-0250160P.  
PR 05-DEC-2000; 2000US-0250391P.  
PR 05-DEC-2000; 2000US-0251030P.  
PR 05-DEC-2000; 2000US-0251988P.  
PR 05-DEC-2000; 2000US-0256719P.  
PR 06-DEC-2000; 2000US-0251479P.  
PR 08-DEC-2000; 2000US-0251856P.  
PR 08-DEC-2000; 2000US-0251868P.  
PR 08-DEC-2000; 2000US-0251869P.  
PR 08-DEC-2000; 2000US-0251989P.  
PR 08-DEC-2000; 2000US-0251990P.

PR 11-DEC-2000; 2000US-0254097P.  
PR 05-JAN-2001; 2001US-0259678P.  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX Rosen CA, Barash SC, Ruben SM;  
XX WPI; 2001-465566/50.  
XX N-PSDB; AAS41545.  
XX Novel polypeptides and polynucleotides useful for diagnosing, preventing,  
XX treating neural, immune system, muscular, reproductive, pulmonary,  
XX cardiovascular, renal, proliferative disorders and cancerous diseases.  
XX Claim 11; SEQ ID NO 1671; 1180pp; English.  
XX The present invention relates to the isolation of novel human enzyme  
XX polypeptides, and the cDNA (AAS40785-AAS41684) and genomic sequences  
XX encoding them. The enzyme polypeptides of the invention may comprise the  
XX functional classes of oxidoreductases, transferases, hydrolases, lyases,  
XX isomerases or ligases. The sequences of the invention are useful in the  
XX diagnosis, treatment, prevention and/or prognosis of a wide range of  
XX disorders including hyperproliferative disorders (e.g. cancer),  
XX immunodeficiency disorders (e.g. AIDS) autoimmune disorders (e.g.  
XX arthritis), neurological disorders (e.g. Alzheimer's disease), metabolic  
XX disorders (e.g. phenylketonuria), inflammatory disorders (e.g. asthma),  
XX cardiovascular disorders (e.g. atherosclerosis), blood-related disorders  
XX (e.g. haemophilia), reproductive disorders (e.g. infertility) and  
XX infectious disorders (e.g. Influenza). The polynucleotides of the  
XX invention can also be used in gene therapy. AAU22915-AAU23814 represent  
XX the novel human enzyme polypeptides of the invention. Note: The sequence  
XX data for this patent did not form part of the printed specification, but  
XX was obtained in electronic format directly from WIPO at  
XX ftp.wipo.int/pub/published\_pct\_sequences  
XX Sequence 309 AA;  
SQ  
Query Match 100.0%; Score 56; DB 4; Length 309;  
Best Local Similarity 100.0%; Pred. No. 0.5;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 IAGIGGEKAGG 11  
Db 48 IAGIGGEKAGG 58  
RESULT 8  
ABG60248  
ID ABG60248 standard; protein; 309 AA.  
XX AC ABG60248;  
XX 13-AUG-2002 (first entry)  
XX Human ovarian antigen #10.  
XX Human; ovarian antigen; ovary disorder; breast disorder;  
XX neoplastic disorder; cancer; infectious disease; inflammatory disease;  
XX reproductive system disorder; autoimmune disorder; Alzheimer's disease;  
XX blood-related disorder; hyperproliferative disorder; hair loss;  
XX urinary system disorder; cardiovascular disorder; arrhythmia;  
XX respiratory disorder; musculoskeletal system disorder;  
XX neural activity disorder; neurological disorder; endocrine disorder;  
XX gastrointestinal disorder; liver disorder; pancreatic disorder;  
XX gall bladder disorder; large intestine disorder; developmental disorder;  
XX inherited disorder; wound healing; skin aging; food additive;  
XX preservative.  
XX Homo sapiens.  
OS  
XX WO200155329-A2.  
XX 02-AUG-2001.  
PD

Qy 1 IAGIGGEKAGG 11  
Db 48 IAGIGGEKAGG 58



RESULT 10  
ABU70735  
ID ABU70735 standard; protein; 339 AA.  
XX  
XX ABU70735;  
XX  
DT 10-JUN-2003 (first entry)

```

XX Human adipocyte Selected Interacting domain, SID, #366.
DE
XX Human; prey; adipocyte; SID; selected interacting domain; anorectic;
XX antidiabetic; protein-protein interaction; diabetes;
KW yeast 2-hybrid assay; metabolic disorder; obesity.
KW
XX Homo sapiens.
OS
XX WO200286122-A2.
PN
XX 31-OCT-2002.
PD
XX 14-MAR-2002; 2002WO-BP003768.
PF
XX 14-MAR-2003; 2001US-0275734P.
PR
XX (HYBR-) HYBRIGENICS.
PA
XX Legrain P, Daviet L;
PI
XX WPI; 2003-103412/09.
XX N-PSDB; ACA57279.
DR
XX New complex between two interacting proteins in adipocyte cells, useful
XX for identifying selected interacting domains that modulate protein
XX interactions, or for preventing or treating metabolic disorders such as
XX obesity or diabetes.
PT
XX Claim 6; Page 229; 382pp; English.
PS
XX The invention relates to a complex between two interacting proteins in
XX adipocyte cells, given in the specification. The proteins are identified
XX by selecting a bait protein from a known adipocyte marker and then
XX performing a yeast 2-hybrid selection to isolate prey proteins encoded by
XX members of an adipocyte cDNA library. The proteins are designated SID
XX (RTM) (selected interacting domains) proteins. Also included are a
XX polynucleotide encoding a polypeptide in the adipocyte cells, a
XX recombinant host cell expressing at least one of the interacting
XX polypeptides of the complex, selecting a modulating compound in adipocyte
XX cells, a SID (RTM) polypeptide comprising any of the 738 amino acid
XX sequences given in the specification (including its fragment or variant),
XX a SID (RTM) polynucleotide comprising any of the 738 nucleotide sequences
XX given in the specification (including its fragment or variant), a vector
XX comprising the SID (RTM) polynucleotide, a recombinant host cell
XX comprising the vector, a protein chip comprising the polypeptides and a
XX record comprising all or part of the data, listed in the specification.
XX The complex, polypeptides, polynucleotides and compounds are useful for
XX preventing or treating metabolic disorders such as obesity or diabetes.
XX The polynucleotides are useful as probes or primers. The complex is
XX particularly useful for identifying selected interacting domains (SID
XX (RTM)) for screening drugs that modulate the protein interaction, thus
XX exhibiting the therapeutic effect. The present sequence represents a SID
XX (prey) protein of the invention
XX
XX Sequence 339 AA;
SQ
Query Match 100.0%; Score 56; DB 6; Length 339;
Best Local Similarity 100.0%; Pred. No. 0.55;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

XX 1 IAGIGGEXAGG 11
DB 78 IAGIGGEXAGG 88
|||||
RESULT 11
AAW12843
ID AAW12843 standard; peptide; 623 AA.
XX
XX AAW12843;
AC
XX 15-DEC-1997 (first entry)
DT

XX Human adipocyte Selected Interacting domain, SID, #366.
DE
XX Human; prey; adipocyte; SID; selected interacting domain; anorectic;
XX antidiabetic; protein-protein interaction; diabetes;
KW yeast 2-hybrid assay; metabolic disorder; obesity.
KW
XX Homo sapiens.
OS
XX WO200286122-A2.
PN
XX 31-OCT-2002.
PD
XX 14-MAR-2002; 2002WO-BP003768.
PF
XX 14-MAR-2003; 2001US-0275734P.
PR
XX (HYBR-) HYBRIGENICS.
PA
XX Legrain P, Daviet L;
PI
XX WPI; 2003-103412/09.
XX N-PSDB; ACA57279.
DR
XX New complex between two interacting proteins in adipocyte cells, useful
XX for identifying selected interacting domains that modulate protein
XX interactions, or for preventing or treating metabolic disorders such as
XX obesity or diabetes.
PT
XX Claim 6; Page 229; 382pp; English.
PS
XX The invention relates to a complex between two interacting proteins in
XX adipocyte cells, given in the specification. The proteins are identified
XX by selecting a bait protein from a known adipocyte marker and then
XX performing a yeast 2-hybrid selection to isolate prey proteins encoded by
XX members of an adipocyte cDNA library. The proteins are designated SID
XX (RTM) (selected interacting domains) proteins. Also included are a
XX polynucleotide encoding a polypeptide in the adipocyte cells, a
XX recombinant host cell expressing at least one of the interacting
XX polypeptides of the complex, selecting a modulating compound in adipocyte
XX cells, a SID (RTM) polypeptide comprising any of the 738 amino acid
XX sequences given in the specification (including its fragment or variant),
XX a SID (RTM) polynucleotide comprising any of the 738 nucleotide sequences
XX given in the specification (including its fragment or variant), a vector
XX comprising the SID (RTM) polynucleotide, a recombinant host cell
XX comprising the vector, a protein chip comprising the polypeptides and a
XX record comprising all or part of the data, listed in the specification.
XX The complex, polypeptides, polynucleotides and compounds are useful for
XX preventing or treating metabolic disorders such as obesity or diabetes.
XX The polynucleotides are useful as probes or primers. The complex is
XX particularly useful for identifying selected interacting domains (SID
XX (RTM)) for screening drugs that modulate the protein interaction, thus
XX exhibiting the therapeutic effect. The present sequence represents a SID
XX (prey) protein of the invention
XX
XX Sequence 339 AA;
SQ
Query Match 100.0%; Score 56; DB 6; Length 339;
Best Local Similarity 100.0%; Pred. No. 0.55;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

XX 1 IAGIGGEXAGG 11
DB 78 IAGIGGEXAGG 88
|||||
RESULT 11
AAW12843
ID AAW12843 standard; peptide; 623 AA.
XX
XX AAW12843;
AC
XX 15-DEC-1997 (first entry)
DT

XX Pro-alpha(III): (I) CP chimeric protein.
DE
XX C-propeptide; recognition sequence; procollagen; monomer chain; therapy;
XX trimerised pro-alpha chain; fibril; procollagen suicide; wound healing;
KW fibrotic disease; human; chimeric protein.
KW
XX Homo sapiens.
OS
XX WO9708311-A1.
PN
XX 06-MAR-1997.
PD
XX 30-AUG-1996; 96WO-GB002122.
PF
XX 31-AUG-1995; 95GB-00017773.
PR
XX 23-MAR-1996; 96GB-00006152.
PR
XX 14-JUN-1996; 96GB-00012476.
XX
XX (UYMA-) UNIV VICTORIA MANCHESTER.
PA
XX Bulleid N, Kadler K;
PI
XX WPI; 1997-179268/16.
XX
XX Novel pro:collagen mol. - comprising pro:collagen C-pro:peptide attached
XX to an alien collagen alpha-chain or non-collagen material, useful e.g.
XX for wound healing.
PT
XX Example 1; Page 32-35; 69pp; English.
PS
XX This sequence represents a chimeric procollagen molecule of the
XX invention. This sequence has the procollagen C-propeptide from the pro
XX -alpha2(I) chain attached to the pro-alpha(III) chain sequence. The C-
XX propeptide is implicated in the assembly of the monomer chains into
XX trimerised pro-alpha chains prior to cleavage of the N- and C-propeptides
XX and formation of collagen in fibril-forming pro-alpha chains. The C-
XX propeptides determine the type-specific assembly of the moieties to which
XX they are attached. The molecule of the invention comprises a first moiety
XX having procollagen C-propeptide activity attached to a second moiety,
XX which is an alien collagen alpha-chain or a non-collagen material. The
XX novel collagen molecule can be used for treatment or diagnosis in humans
XX or animals, especially for the treatment of procollagen suicide, as an
XX adhesive or implant, to promote (chronic) wound healing or fibrotic
XX diseases with reduced scarring or for use in photography, brewing,
XX foodstuffs or textiles. The novel collagen molecules, especially when
XX containing substitutions in the recognition site, may have significantly
XX altered properties and characteristics, such as different binding
XX kinetics or alpha-chain selection properties
XX
XX Sequence 623 AA;
SQ
Query Match 100.0%; Score 56; DB 2; Length 623;
Best Local Similarity 100.0%; Pred. No. 1;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

XX 1 IAGIGGEXAGG 11
DB 365 IAGIGGEXAGG 375
|||||
RESULT 12
AAW12842
ID AAW12842 standard; protein; 626 AA.
XX
XX AAW12842;
AC
XX 15-DEC-1997 (first entry)
DT
XX Truncated pro-alpha(III) chain.
XX
XX C-propeptide; recognition sequence; procollagen; monomer chain; therapy;
XX trimerised pro-alpha chain; fibril; procollagen suicide; wound healing;
KW

```

KW fibrotic disease; human.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO9708311-A1.  
 XX  
 PD 06-MAR-1997.  
 XX  
 PF 30-AUG-1996; 96WO-GB0002122.  
 XX  
 PR 31-AUG-1995; 95GB-00017773.  
 PR 23-MAR-1996; 96GB-00006152.  
 PR 14-JUN-1996; 96GB-00012476.  
 XX  
 PA (UYMA-) UNIV VICTORIA MANCHESTER.  
 XX  
 XX  
 PI Bulleid N, Kadler K;  
 XX  
 XX WPI; 1997-179268/16.  
 DR N-PSDB; AAT59892.  
 DR  
 XX  
 XX Novel pro:collagen mol. - comprising pro:collagen C-pro:peptide attached  
 PT to an alien collagen alpha-chain or non-collagen material, useful e.g.  
 PT for wound healing.  
 PT  
 XX  
 XX Example 1; Page 28-31; 69pp; English.  
 PS  
 XX This sequence represents a truncated procollagen pro-alpha(III) chain  
 CC that can be used in the procollagen molecules of the invention. The C-  
 CC propeptide is implicated in the assembly of the monomer chains into  
 CC trimerised pro-alpha chains prior to cleavage of the N- and C-propeptides  
 CC and formation of collagen in fibril-forming pro-alpha chains. The C-  
 CC propeptides determine the type-specific assembly of the moieties to which  
 CC they are attached. The molecule of the invention comprises a first moiety  
 CC having procollagen C-propeptide activity attached to a second moiety,  
 CC which is an alien collagen alpha-chain or a non-collagen material. The  
 CC novel collagen molecule can be used for treatment or diagnosis in humans  
 CC or animals, especially for the treatment of procollagen suicide, as an  
 CC adhesive or implant, to promote (chronic) wound healing or fibrotic  
 CC diseases with reduced scarring or for use in photography, brewing,  
 CC foodstuffs or textiles. The novel collagen molecules, especially when  
 CC containing substitutions in the recognition site, may have significantly  
 CC altered properties and characteristics, such as different binding  
 CC kinetics or alpha-chain selection properties  
 XX  
 SQ Sequence 626 AA;  
 Query Match 100.0%; Score 56; DB 2; Length 626;  
 Best Local Similarity 100.0%; Pred. No. 1;  
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 IAGIGGEKAGG 11  
 Db 365 IAGIGGEKAGG 375  
 DE  
 XX  
 XX ADE08475 standard; protein; 842 AA.  
 AC ADE08475;  
 XX  
 XX 29-JAN-2004 (first entry)  
 DT  
 XX Novel protein (useful for identifying genetic disorders) #630.  
 DE  
 XX novel gene; novel protein; tissue marker; molecular weight marker;  
 KW chromosome marker; genetic disorder.  
 KW  
 XX Unidentified.  
 OS  
 XX  
 XX WO2003054152-A2.  
 PN  
 XX

PD 03-JUL-2003.  
 XX  
 XX 10-DEC-2002; 2002WO-US039555.  
 XX  
 PR 10-DEC-2001; 2001US-0339739P.  
 PR 11-DEC-2001; 2001US-0339453P.  
 PR 14-MAR-2002; 2002US-0365091P.  
 PR 12-MAR-2002; 2002US-0365384P.  
 PR 12-APR-2002; 2002US-0372381P.  
 PR 12-APR-2002; 2002US-0372615P.  
 PR 22-APR-2002; 2002US-00128558.  
 PR 24-APR-2002; 2002US-0376045P.  
 XX  
 PA (HYSE-) HYSEQ INC.  
 XX  
 XX Tang YT, Asundi V, Goodrich RW, Ren F, Zhang J, Zhao QA, Wang J;  
 PI Ghosh M, Xue AJ, Wehrman T, Weng G, Zhou P, Drmanac RT, Wang Z;  
 PI Ma Y, Wang D, Chen R, Xu C, Boyle BJ;  
 XX  
 DR WPI; 2003-569235/53.  
 DR N-PSDB; ADE07564.  
 DR  
 XX New polynucleotides, useful for expressing recombinant proteins for  
 PT analysis, characterization or therapeutic use, or as markers for tissues  
 PT in which the corresponding protein is preferentially expressed.  
 XX  
 PS Claim 20; SEQ ID NO 1541; 1177pp; English.  
 XX  
 XX The invention comprises the amino acid and coding sequences of novel  
 CC proteins. The DNA and protein sequences of the invention are useful as:  
 CC markers for tissues in which the corresponding protein is preferentially  
 CC expressed; as molecular weight markers on gels; as chromosome markers or  
 CC tags; to identify chromosomes or to map related gene positions; and to  
 CC compare with endogenous DNA sequences in patients to identify potential  
 CC genetic disorders. The present amino acid sequence represents a protein  
 CC of the invention.  
 XX  
 SQ Sequence 842 AA;  
 Query Match 100.0%; Score 56; DB 7; Length 842;  
 Best Local Similarity 100.0%; Pred. No. 1.4;  
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 IAGIGGEKAGG 11  
 Db 581 IAGIGGEKAGG 591  
 DE  
 XX  
 XX ABB50291 standard; protein; 1466 AA.  
 AC ABB50291;  
 XX  
 XX 08-FEB-2002 (first entry)  
 DT  
 XX Collagen type III alpha-1 ovarian tumour marker protein, SEQ ID NO:72.  
 DE  
 XX  
 XX Ovarian tumour marker gene; human; overexpression; upregulation;  
 KW epithelial tumour; cancer; diagnosis; prognosis; disease monitoring;  
 KW identification; serous cystadenoma; borderline serous tumour;  
 KW serous cystadenocarcinoma; mucinous cystadenocarcinoma;  
 KW mucinous cystadenoma; borderline mucinous tumour; endometrioid carcinoma;  
 KW undifferentiated carcinoma; clear cell adenocarcinoma; cystadenofibroma;  
 KW adenofibroma; Brenner tumour; serial analysis of gene expression; SAGE;  
 KW immune response pathway; cell proliferation regulation; protein folding;  
 KW membrane localised; secreted; therapeutic target; cytostatic;  
 KW gene therapy; vaccine.  
 XX  
 OS Homo sapiens.  
 XX  
 XX WO200175177-A2.  
 PN  
 XX

Mon Sep 20 11:05:46 2004

peoriasis.  
 KW XX OS Homo sapiens.  
 XX XX WO200210217-A2.  
 XX XX 07-FEB-2002.  
 XX XX 01-AUG-2001; 2001WO-US024031.  
 XX XX 02-AUG-2000; 2000US-0222599P.  
 XX XX 11-AUG-2000; 2000US-0224360P.  
 XX XX 11-APR-2001; 2001US-0282850P.  
 XX XX (UYJO ) UNIV JOHNS HOPKINS.  
 XX XX St Croix B, Kinzler KW, Vogelstein B;  
 XX XX N-PSDB; ABL92101.  
 XX XX WPI; 2002-291856/33.  
 XX XX N-PSDB; ABL92101.  
 XX XX An isolated molecule comprising an antibody variable region which  
 PT specifically binds to an extracellular domain of a tumor endothelial  
 PT marker (TEM) protein, useful for inhibiting tumor growth.  
 XX Claim 68; Page 197-200; 331pp; English.  
 XX The invention relates to an isolated molecule comprising an antibody  
 CC variable region which specifically binds to an extracellular domain of a  
 CC tumor endothelial marker (TEM) protein selected from ABB90732, ABB90740,  
 CC ABB90749, ABB90750 and ABB90769. The antibodies which bind to TEM  
 CC proteins have cytostatic, immunostimulant and antiangiogenic activity.  
 CC They are useful for inhibiting tumor growth, neovascularization in subjects  
 CC bearing a vascularised tumour, polycystic kidney disease, diabetic  
 CC retinopathy, rheumatoid arthritis and psoriasis. Human, mouse and rat TEM  
 CC genes and the encoded proteins (ABL92075-ABL92141 and ABB90721-ABB90789)  
 CC are disclosed, as are marker oligonucleotide sequences: tumour  
 CC endothelial markers (TEM) ABL91996-ABL92041 and ABL92143-ABL92191; normal  
 CC endothelial markers (NEM) ABL92042-ABL92074; and pan-endothelial markers  
 CC (PEM) ABL91903-ABL91995  
 XX Sequence 1466 AA;  
 SQ Query Match 100.0%; Score 56; DB 5; Length 1466;  
 Best Local Similarity 100.0%; Pred. No. 2.5;  
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 1 IAGIGGKAGG 11  
 Db 1205 IAGIGGKAGG 1215

Search completed: September 18, 2004, 04:25:32  
 Job time : 58.8824 secs

11-OCT-2001.  
 PD XX 03-APR-2001; 2001WO-US010947.  
 PF XX 03-APR-2000; 2000US-0194336P.  
 XX XX (USSH ) US DEPT HEALTH & HUMAN SERVICES.  
 XX XX Morin PJ, Sherman-Baust CA, Pizer ES, Hough CD;  
 XX XX WPI; 2001-626450/72.  
 XX XX N-PSDB; ABA83117.  
 XX XX Detecting and identifying ovarian tumor, identifying increased risk for  
 PT developing ovarian cancer, and determining effectiveness of ovarian  
 PT cancer treatment, by measuring expression level of ovarian tumor marker  
 PT gene.  
 XX Claim 23; Page 114-117; 140pp; English.  
 XX The invention relates to methods for diagnosing and prognosing ovarian  
 CC tumours in an individual via the detection and measurement of the  
 CC expression of ovarian tumor marker genes (ABA83081-ABA83122, ABA83180,  
 CC ABA83182 and ABA83184) or segments thereof (ABA83123-ABA83169, ABA83179,  
 CC ABA83181 and ABA83183). The methods of the invention are useful for  
 CC detecting an ovarian tumor in a patient, for identifying an individual  
 CC at increased risk for developing ovarian cancer, in prognostic tests for  
 CC assessing the relative severity of ovarian cancer, in tests for  
 CC monitoring a patient in remission from ovarian cancer and in tests for  
 CC monitoring disease status in a patient being treated for ovarian cancer.  
 CC The methods can additionally be used to identify a particular tumour as  
 CC being an ovarian tumour (i.e., an epithelial ovarian tumour selected from  
 CC serous cystadenoma, borderline serous tumour, serous cystadenocarcinoma,  
 CC mucinous cystadenoma, borderline mucinous tumour, mucinous  
 CC cystadenocarcinoma, endometrioid carcinoma, undifferentiated carcinoma,  
 CC clear cell adenocarcinoma, cystadenofibroma, adenofibroma and Brenner  
 CC tumour. The ovarian tumour marker genes of the invention were identified  
 CC using SAGE (serial analysis of gene expression) and were found to be  
 CC overexpressed in a broad variety of ovarian epithelial tumour cells  
 CC relative to normal ovarian epithelial cells. The marker genes are  
 CC implicated in immune response pathways, in the regulation of cell  
 CC proliferation and in protein folding, and many of these are membrane-  
 CC localised or secreted. In addition to their use as diagnostic and  
 CC prognostic markers, the ovarian tumour marker genes or their encoded  
 CC proteins may be used as therapeutic targets for the treatment and  
 CC prevention of ovarian cancer. Sequences ABB50257-ABB50299 represent  
 CC proteins encoded by ovarian tumour marker genes of the invention  
 XX Sequence 1466 AA;  
 SQ Query Match 100.0%; Score 56; DB 4; Length 1466;  
 Best Local Similarity 100.0%; Pred. No. 2.5;  
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 1 IAGIGGKAGG 11  
 Db 1205 IAGIGGKAGG 1215

RESULT 15  
 ABB90747  
 ID ABB90747 standard; protein; 1466 AA.  
 XX XX ABB90747;  
 XX XX 30-MAY-2002 (first entry)  
 XX XX Human Tumour Endothelial Marker polypeptide SEQ ID NO 226.  
 XX Human; mouse; rat; TEM; tumour endothelial marker; NEM; PEM; cytostatic;  
 KW normal endothelial marker; pan-endothelial marker; immunostimulant;  
 KW antiangiogenic; tumour; neovascularization; vascularised tumour;  
 KW polycystic kidney disease; diabetes; retinopathy; rheumatoid arthritis;

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 18, 2004, 04:20:54 ; Search time 18.1176 Seconds  
(without alignments)  
31.344 Million cell updates/sec

Title: US-10-615-959-42  
Perfect score: 56  
Sequence: 1 IAGIGGEKAGG 11

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA:\*

1: /cgn2\_6/ptodata/2/iaa/5A COMB.pep.\*  
2: /cgn2\_6/ptodata/2/iaa/5B COMB.pep.\*  
3: /cgn2\_6/ptodata/2/iaa/6A COMB.pep.\*  
4: /cgn2\_6/ptodata/2/iaa/6B COMB.pep.\*  
5: /cgn2\_6/ptodata/2/iaa/PTUS COMB.pep.\*  
6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	56	100.0	11	4	US-10-009-999A-42
2	56	100.0	12	4	US-10-009-999A-43
3	56	100.0	623	3	US-09-029-348-3
4	56	100.0	626	3	US-09-029-348-2
5	50	89.3	1078	3	US-08-963-825-21
6	50	89.3	1078	4	US-09-500-811-21
7	50	89.3	1078	4	US-09-570-573-21
8	50	89.3	1078	4	US-09-548-608-21
9	48	85.7	15	4	US-08-278-774-20
10	42	75.0	8	4	US-10-009-999A-40
11	42	75.0	9	4	US-10-009-999A-41
12	42	75.0	11	6	5473052-20
13	41	73.2	221	4	US-09-252-991A-28805
14	41	73.2	503	4	US-09-252-991A-23643
15	40	71.4	674	4	US-09-252-991A-26476
16	39	69.6	157	2	US-08-795-868-4
17	39	69.6	247	4	US-09-199-637A-363
18	39	69.6	247	4	US-09-252-991A-21412
19	39	69.6	591	4	US-09-489-033A-12837
20	39	69.6	674	4	US-09-199-637A-51
21	38	67.9	227	4	US-09-252-991A-30541
22	38	67.9	419	4	US-09-543-681A-7295
23	38	67.9	490	4	US-09-252-991A-22715
24	38	67.9	576	4	US-09-252-991A-32499
25	38	67.9	968	4	US-09-252-991A-18925
26	37	66.1	227	4	US-09-252-991A-30027
27	37	66.1	310	4	US-09-252-991A-19986

28	37	66.1	336	4	US-09-252-991A-21790	Sequence 21790, A
29	37	66.1	811	4	US-09-252-991A-28570	Sequence 28570, A
30	37	66.1	927	4	US-09-252-991A-16765	Sequence 16765, A
31	36	64.3	75	4	US-08-924-629C-17	Sequence 17, Appl
32	36	64.3	75	4	US-08-924-629C-18	Sequence 18, Appl
33	36	64.3	144	2	US-08-997-080-54	Sequence 54, Appl
34	36	64.3	144	2	US-08-873-970-54	Sequence 54, Appl
35	36	64.3	144	3	US-09-095-855-54	Sequence 54, Appl
36	36	64.3	144	3	US-08-705-347A-54	Sequence 54, Appl
37	36	64.3	144	3	US-09-324-542-54	Sequence 54, Appl
38	36	64.3	144	4	US-09-205-426-54	Sequence 54, Appl
39	36	64.3	144	4	US-09-200-643-54	Sequence 54, Appl
40	36	64.3	345	4	US-09-252-991A-18076	Sequence 18076, A
41	36	64.3	412	4	US-09-252-991A-29886	Sequence 29886, A
42	36	64.3	460	4	US-09-056-556-184	Sequence 184, App
43	36	64.3	460	4	US-09-072-596-179	Sequence 179, App
44	36	64.3	460	4	US-09-072-967-184	Sequence 184, App
45	36	64.3	460	4		

ALIGNMENTS

RESULT 1

US-10-009-999A-42  
; Sequence 42, Application US/100099999A  
; Patent No. 6602980  
; GENERAL INFORMATION:  
; APPLICANT: Eyre, David R.  
; TITLE OF INVENTION: SYNTHETIC PEPTIDES FOR COLLAGEN RESORPTION ASSAYS  
; FILE REFERENCE: WROS-1-18220  
; CURRENT APPLICATION NUMBER: US/10/009,999A  
; PRIOR FILING DATE: 2003-01-22  
; PRIOR APPLICATION NUMBER: PCT/US99/29357  
; PRIOR FILING DATE: 1999-12-10  
; PRIOR APPLICATION NUMBER: US 09/335,098  
; PRIOR FILING DATE: 1999-06-17  
; PRIOR APPLICATION NUMBER: US 60/141,574  
; PRIOR FILING DATE: 1999-06-29  
; PRIOR APPLICATION NUMBER: US 60/142,274  
; PRIOR FILING DATE: 1999-07-02  
; PRIOR APPLICATION NUMBER: US 60/142,675  
; PRIOR FILING DATE: 1999-07-07  
; PRIOR APPLICATION NUMBER: US 09/385,740  
; PRIOR FILING DATE: 1999-08-30  
; NUMBER OF SEQ ID NOS: 45  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 42  
; LENGTH: 11  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: MISC FEATURE  
; OTHER INFORMATION: synthetic peptide corresponding to carboxy-terminal  
; OTHER INFORMATION: telopeptide sequence of human type III collagen  
US-10-009-999A-42

Query Match 100.0%; Score 56; DB 4; Length 11;  
Best Local Similarity 100.0%; Pred. No. 0.0077;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IAGIGGEKAGG 11

DB 1 IAGIGGEKAGG 11

RESULT 2

US-10-009-999A-43  
; Sequence 43, Application US/100099999A  
; Patent No. 6602980  
; GENERAL INFORMATION:  
; APPLICANT: Eyre, David R.  
; TITLE OF INVENTION: SYNTHETIC PEPTIDES FOR COLLAGEN RESORPTION ASSAYS

```
; FILE REFERENCE: WROS-1-18220
; CURRENT APPLICATION NUMBER: US/10/009,999A
; CURRENT FILING DATE: 2003-01-22
; PRIOR APPLICATION NUMBER: PCT/US99/29357
; PRIOR FILING DATE: 1999-12-10
; PRIOR APPLICATION NUMBER: US 09/335,098
; PRIOR FILING DATE: 1999-06-17
; PRIOR APPLICATION NUMBER: US 60/141,574
; PRIOR FILING DATE: 1999-06-29
; PRIOR APPLICATION NUMBER: US 60/142,274
; PRIOR FILING DATE: 1999-07-02
; PRIOR APPLICATION NUMBER: US 60/142,675
; PRIOR FILING DATE: 1999-07-07
; PRIOR APPLICATION NUMBER: US 09/385,740
; PRIOR FILING DATE: 1999-08-30
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 43
; LENGTH: 12
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: MISC FEATURE
; OTHER INFORMATION: synthetic peptide corresponding to carboxy-terminal
; OTHER INFORMATION: telopeptide sequence of human type III collagen
US-10-009-999A-43
```

```
Query Match 100.0%; Score 56; DB 4; Length 12;
Best Local Similarity 100.0%; Pred. No. 0.0084;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 IAGIGGEKAGG 11
Db 1 IAGIGGEKAGG 11
```

```
RESULT 3
US-09-029-348-3
; Sequence 3, Application US/09029348
; Patent No. 6171827
; GENERAL INFORMATION:
; APPLICANT: THE VICTORIA UNIVERSITY OF MANCHESTER
; TITLE OF INVENTION: NOVEL PROCOLLAGENS
; FILE REFERENCE: d087857PUS LISTING
; CURRENT APPLICATION NUMBER: US/09/029,348
; CURRENT FILING DATE: 1998-05-07
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 623
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: SEQUENCE
; OTHER INFORMATION: DERIVED FROM cDNA OF PROCOLLAGENS
US-09-029-348-3
```

```
Query Match 100.0%; Score 56; DB 3; Length 623;
Best Local Similarity 100.0%; Pred. No. 0.37;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 IAGIGGEKAGG 11
Db 365 IAGIGGEKAGG 375
```

```
RESULT 4
US-09-029-348-2
; Sequence 2, Application US/09029348
; Patent No. 6171827
; GENERAL INFORMATION:
; APPLICANT: THE VICTORIA UNIVERSITY OF MANCHESTER
; TITLE OF INVENTION: NOVEL PROCOLLAGENS
```

```
; FILE REFERENCE: d087857PUS LISTING
; CURRENT APPLICATION NUMBER: US/09/029,348
; CURRENT FILING DATE: 1998-05-07
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 626
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: SEQUENCE
; OTHER INFORMATION: DERIVED FROM cDNA OF PROCOLLAGENS
US-09-029-348-2
```

```
Query Match 100.0%; Score 56; DB 3; Length 626;
Best Local Similarity 100.0%; Pred. No. 0.37;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 IAGIGGEKAGG 11
Db 365 IAGIGGEKAGG 375
```

```
RESULT 5
US-08-963-825-21
; Sequence 21, Application US/08963825
; Patent No. 6110689
; GENERAL INFORMATION:
; APPLICANT: Ovist, Per
; APPLICANT: Bonde, Martin
; TITLE OF INVENTION: A Method for Assaying Collagen Fragments
; TITLE OF INVENTION: In Body Fluids, A Test Kit and Means for Carrying Out the
; TITLE OF INVENTION: Method and Use of the Method to Diagnose the Presence of
; TITLE OF INVENTION: Disorders Associated with the Metabolism of
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Darby & Darby PC
; STREET: 805 Third Avenue
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/963,825
; FILING DATE:
; CLASSIFICATION: 436
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/187,319
; FILING DATE: 21-JAN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Gogoris, Adda C
; REGISTRATION NUMBER: 29,714
; REFERENCE/DOCKET NUMBER: 4305/08701
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-527-7700
; TELEFAX: 212-753-6237
; TELEX: 236687
; INFORMATION FOR SEQ ID NO: 21:
```

```
SEQUENCE CHARACTERISTICS:
LENGTH: 1078 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
IMMEDIATE SOURCE:
CLONE: COLLAGEN ALPHA 1 (III)
US-08-963-825-21
```

```

Query Match      89.3%; Score 50; DB 3; Length 1078;
Best Local Similarity 90.9%; Pred. No. 5.3;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 IAGIGGEKAGG 11
    ||||| |||||
Db 1058 IAGIGAEGKAGG 1068

RESULT 6
US-09-500-811-21
; Sequence 21, Application US/09500811
; Patent No. 6323314
; GENERAL INFORMATION:
; APPLICANT: Qvist, Per
; APPLICANT: Bonde, Martin
; TITLE OF INVENTION: A Method for Assaying Collagen Fragments
; TITLE OF INVENTION: in Body Fluids, A Test Kit and Means for Carrying Out the
; TITLE OF INVENTION: Method and Use of the Method to Diagnose the Presence of
; TITLE OF INVENTION: Disorders Associated with the Metabolism of
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Darby & Darby PC
; STREET: 805 Third Avenue
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/500,811
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA: 08/187,319
; APPLICATION NUMBER: 08/187,319
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Gogoris, Adda C
; REGISTRATION NUMBER: 29,714
; REFERENCE/DOCKET NUMBER: 4305/08701
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-527-7700
; TELEFAX: 212-753-6237
; TELEX: 236687
; INFORMATION FOR SEQ ID NO: 21:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1078 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; IMMEDIATE SOURCE:
; CLONE: COLLAGEN ALPHA 1 (III)
; US-09-500-811-21

Query Match      89.3%; Score 50; DB 4; Length 1078;
Best Local Similarity 90.9%; Pred. No. 5.3;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 IAGIGGEKAGG 11
    ||||| |||||
Db 1058 IAGIGAEGKAGG 1068

RESULT 7
US-09-570-573-21
; Sequence 21, Application US/09570573

```

```

; Patent No. 6342361
; GENERAL INFORMATION:
; APPLICANT: Qvist, Per
; APPLICANT: Bonde, Martin
; TITLE OF INVENTION: A Method for Assaying Collagen Fragments
; TITLE OF INVENTION: in Body Fluids, A Test Kit and Means for Carrying Out the
; TITLE OF INVENTION: Method and Use of the Method to Diagnose the Presence of
; TITLE OF INVENTION: Disorders Associated with the Metabolism of
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Darby & Darby PC
; STREET: 805 Third Avenue
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/570,573
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA: 08/187,319
; APPLICATION NUMBER: 08/187,319
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Gogoris, Adda C
; REGISTRATION NUMBER: 29,714
; REFERENCE/DOCKET NUMBER: 4305/08701
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-527-7700
; TELEFAX: 212-753-6237
; TELEX: 236687
; INFORMATION FOR SEQ ID NO: 21:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1078 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; IMMEDIATE SOURCE:
; CLONE: COLLAGEN ALPHA 1 (III)
; US-09-570-573-21

Query Match      89.3%; Score 50; DB 4; Length 1078;
Best Local Similarity 90.9%; Pred. No. 5.3;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 IAGIGGEKAGG 11
    ||||| |||||
Db 1058 IAGIGAEGKAGG 1068

RESULT 8
US-09-548-608-21
; Sequence 21, Application US/09548608
; Patent No. 6355442
; GENERAL INFORMATION:
; APPLICANT: Qvist, Per
; APPLICANT: Bonde, Martin
; TITLE OF INVENTION: A Method for Assaying Collagen Fragments
; TITLE OF INVENTION: in Body Fluids, A Test Kit and Means for Carrying Out the
; TITLE OF INVENTION: Method and Use of the Method to Diagnose the Presence of
; TITLE OF INVENTION: Disorders Associated with the Metabolism of
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Darby & Darby PC
; STREET: 805 Third Avenue
; CITY: New York

```

STATE: New York  
COUNTRY: USA  
ZIP: 10022  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/548,608  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/187,319  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Gogoris, Adda C  
REGISTRATION NUMBER: 29,714  
REFERENCE/DOCKET NUMBER: 4305/08701  
TELEPHONE: 212-527-7700  
TELEFAX: 212-753-6237  
TELEX: 236687  
INFORMATION FOR SEQ ID NO: 21:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1078 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
ORIGINAL SOURCE:  
ORGANISM: Homo sapiens  
IMMEDIATE SOURCE:  
CLONE: COLLAGEN ALPHA 1 (III)  
US-09-548-608-21

Query Match 89.3%; Score 50; DB 4; Length 1078;  
Best Local Similarity 90.9%; Pred. No. 5.3;  
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 IAGIGGEKAGG 11  
Db 1058 IAGIGAERKAGG 1068

RESULT 9  
US-08-278-774-20  
Sequence 20, Application US/08278774  
Patent No. 6653450  
GENERAL INFORMATION:  
APPLICANT: Berg, Richard A  
APPLICANT: Toman, David P  
APPLICANT: Wallace, Donald  
TITLE OF INVENTION: MUTATED RECOMBINANT COLLAGENS  
NUMBER OF SEQUENCES: 33  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: COLLAGEN CORPORATION  
STREET: 2500 Faber Place  
City: Palo Alto  
STATE: California  
COUNTRY: USA  
ZIP: 94303  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/278,774  
FILING DATE: 22-JUL-1994  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Rafayko, Kathi L  
REGISTRATION NUMBER: 36,644

REFERENCE/DOCKET NUMBER: 94-018  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 354-4642  
TELEFAX: (415) 354-4752  
TELEX:  
INFORMATION FOR SEQ ID NO: 20:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 15 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-278-774-20

Query Match 85.7%; Score 48; DB 4; Length 15;  
Best Local Similarity 100.0%; Pred. No. 0.18;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 GIGGEKAGG 11  
Db 1 GIGGEKAGG 9

RESULT 10  
US-10-009-999A-40  
Sequence 40, Application US/100099999A  
Patent No. 6602980  
GENERAL INFORMATION:  
APPLICANT: Eyre, David R.  
TITLE OF INVENTION: SYNTHETIC PEPTIDES FOR COLLAGEN RESORPTION ASSAYS  
FILE REFERENCE: WROS-1-18220  
CURRENT APPLICATION NUMBER: US/10/009,999A  
CURRENT FILING DATE: 2003-01-22  
PRIOR APPLICATION NUMBER: PCT/US99/29357  
PRIOR FILING DATE: 1999-12-10  
PRIOR APPLICATION NUMBER: US 09/335,098  
PRIOR FILING DATE: 1999-06-17  
PRIOR APPLICATION NUMBER: US 60/141,574  
PRIOR FILING DATE: 1999-06-29  
PRIOR APPLICATION NUMBER: US 60/142,274  
PRIOR FILING DATE: 1999-07-02  
PRIOR APPLICATION NUMBER: US 60/142,675  
PRIOR FILING DATE: 1999-07-07  
PRIOR APPLICATION NUMBER: US 09/385,740  
PRIOR FILING DATE: 1999-08-30  
NUMBER OF SEQ ID NOS: 45  
SOFTWARE: PatentIn version 3.2  
SEQ ID NO 40  
LENGTH: 8  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: MISC FEATURE  
OTHER INFORMATION: synthetic peptide corresponding to carboxy-terminal  
OTHER INFORMATION: telopeptide sequence of human type III collagen  
US-10-009-999A-40

Query Match 75.0%; Score 42; DB 4; Length 8;  
Best Local Similarity 100.0%; Pred. No. 3e+05;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 IGGEKAGG 11  
Db 1 IGGEKAGG 8

RESULT 11  
US-10-009-999A-41  
Sequence 41, Application US/100099999A  
Patent No. 6602980  
GENERAL INFORMATION:  
APPLICANT: Eyre, David R.  
TITLE OF INVENTION: SYNTHETIC PEPTIDES FOR COLLAGEN RESORPTION ASSAYS



```
; FILE REFERENCE: WR0S-1-18220
; CURRENT APPLICATION NUMBER: US/10/009,999A
; CURRENT FILING DATE: 2003-01-22
; PRIOR APPLICATION NUMBER: PCT/US99/29357
; PRIOR FILING DATE: 1999-12-10
; PRIOR APPLICATION NUMBER: US 09/335,098
; PRIOR FILING DATE: 1999-06-17
; PRIOR APPLICATION NUMBER: US 60/141,574
; PRIOR FILING DATE: 1999-06-29
; PRIOR APPLICATION NUMBER: US 60/142,274
; PRIOR FILING DATE: 1999-07-02
; PRIOR APPLICATION NUMBER: US 60/142,675
; PRIOR FILING DATE: 1999-07-07
; PRIOR APPLICATION NUMBER: US 09/385,740
; PRIOR FILING DATE: 1999-08-30
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 41
; LENGTH: 9
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: MISC_FEATURE
; OTHER INFORMATION: synthetic peptide corresponding to carboxy-terminal
; OTHER INFORMATION: telopeptide sequence of human type III collagen
US-10-009-999A-41

Query Match 75.0%; Score 42; DB 4; Length 9;
Best Local Similarity 100.0%; Pred. No. 3e+05; 0; Indels 0; Gaps 0;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 IGGEKAGG 11
   |||||
Db 1 IGGEKAGG 8

RESULT 12
5473052-20
; APPLICANT: EYRE, DAVID R.
; TITLE OF INVENTION: ANTIGEN-BINDING FRAGMENTS OF AN ANTIBODY
; TO TYPE-I COLLAGEN AMINO-TERMINAL TELOPEPTIDE
; NUMBER OF SEQUENCES: 30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/221,705
; FILING DATE: 01-APR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 614,719
; FILING DATE: 21-NOV-1990
; APPLICATION NUMBER: 444,881
; FILING DATE: 01-DEC-1989
; APPLICATION NUMBER: 118,234
; FILING DATE: 06-NOV-1987
; SEQ ID NO:20:
; LENGTH: 11
5473052-20

Query Match 75.0%; Score 42; DB 6; Length 11;
Best Local Similarity 88.9%; Pred. No. 1.1;
Matches 8; Conservative 0; Mismatches 1; Indels 1; Gaps 0;

QY 3 GIGGEKAGG 11
   |||||
Db 1 GIGGEKAGG 9

RESULT 13
US-09-252-991A-28805
; Sequence 28805, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 26476
; LENGTH: 674

Query Match 73.2%; Score 41; DB 4; Length 221;
Best Local Similarity 63.6%; Pred. No. 28;
Matches 7; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 IAGIGGEKAGG 11
   ||:||||:
Db 186 IAAAGGEGSGG 196

RESULT 14
US-09-252-991A-23643
; Sequence 23643, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 23643
; LENGTH: 503
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
; US-09-252-991A-23643

Query Match 73.2%; Score 41; DB 4; Length 503;
Best Local Similarity 70.0%; Pred. No. 62;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 AGIGGEKAGG 11
   ||:||||
Db 481 AGLAGERAGG 490

RESULT 15
US-09-252-991A-26476
; Sequence 26476, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 26476
; LENGTH: 674
```

```

; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-26476

Query Match      71.4%; Score 40; DB 4; Length 674;
Best Local Similarity 63.6%; Pred. No. 1.2e+02;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY      1 IAGIGGKAGG 11
      |::|||
Db      410 ITWGGDKAGG 420

```

Search completed: September 18, 2004, 04:32:23  
 Job time : 19.1176 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 18, 2004, 04:29:56 ; Search time 52.1961 Seconds  
(without alignments)  
67.677 Million cell updates/sec

Title: US-10-615-959-42  
Perfect score: 56  
Sequence: 1 IAGTGGKAGG 11

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1342398 seqs, 321133274 residues

Total number of hits satisfying chosen parameters: 1342398

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications AA.\*  
1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*  
2: /cgn2\_6/ptodata/2/pubpaa/PTCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/2/pubpaa/PTCTUS\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*  
8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/2/pubpaa/US09A\_PUBCOMB.pep.\*  
10: /cgn2\_6/ptodata/2/pubpaa/US09B\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/2/pubpaa/US09C\_PUBCOMB.pep.\*  
12: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep.\*  
13: /cgn2\_6/ptodata/2/pubpaa/US10A\_PUBCOMB.pep.\*  
14: /cgn2\_6/ptodata/2/pubpaa/US10B\_PUBCOMB.pep.\*  
15: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pep.\*  
16: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep.\*  
17: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*  
18: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	56	100.0	11	12	US-10-615-959-42
2	56	100.0	12	12	US-10-615-959-43
3	56	100.0	25	15	US-10-366-125-11
4	56	100.0	262	15	US-10-366-125-19
5	56	100.0	309	9	US-09-908-711-78
6	56	100.0	1466	12	US-09-918-715-226
7	56	100.0	1466	12	US-10-257-021-72
8	56	100.0	1466	14	US-10-177-293-68
9	56	100.0	1466	14	US-10-301-822-33
10	56	100.0	1466	16	US-10-357-851-3
11	56	100.0	1466	16	US-10-358-024-3
12	56	100.0	1466	16	US-10-734-564-103
13	55	98.2	1466	15	US-10-402-089-12
14	55	98.2	1466	15	US-10-402-072A-12
15	50	89.3	1078	14	US-10-058-124-21

16	49	87.5	1466	15	US-10-402-089-4	Sequence 4, Appli
17	49	87.5	1466	15	US-10-402-089-6	Sequence 6, Appli
18	49	87.5	1466	15	US-10-402-072A-4	Sequence 6, Appli
19	49	87.5	1466	15	US-10-402-072A-6	Sequence 6, Appli
20	43	76.8	225	16	US-10-437-963-106003	Sequence 106003,
21	42	75.0	8	12	US-10-615-959-40	Sequence 40, Appl
22	42	75.0	9	12	US-10-615-959-41	Sequence 41, Appl
23	42	75.0	80	16	US-10-437-963-196331	Sequence 196331,
24	42	75.0	112	16	US-10-437-963-139908	Sequence 139908,
25	41	73.2	49	16	US-10-437-963-163129	Sequence 163129,
26	41	73.2	105	16	US-10-767-701-37543	Sequence 37543, A
27	41	73.2	198	16	US-10-437-963-114393	Sequence 114393, A
28	41	73.2	238	12	US-10-425-114-37899	Sequence 37899, A
29	40	71.4	105	16	US-10-767-701-36450	Sequence 36450, A
30	40	71.4	107	12	US-10-425-114-37950	Sequence 37950, A
31	40	71.4	284	12	US-10-437-963-128057	Sequence 128057,
32	40	71.4	300	12	US-10-425-114-47773	Sequence 47773, A
33	40	71.4	304	12	US-10-425-114-61720	Sequence 61720, A
34	40	71.4	467	12	US-10-425-114-70808	Sequence 70808, A
35	40	71.4	471	12	US-10-425-114-64617	Sequence 64617, A
36	40	71.4	696	16	US-10-437-963-189506	Sequence 189506,
37	39	69.6	59	12	US-10-424-599-253049	Sequence 253049,
38	39	69.6	117	12	US-10-425-114-57398	Sequence 57398, A
39	39	69.6	137	16	US-10-437-963-114626	Sequence 114626,
40	39	69.6	200	16	US-10-437-963-180655	Sequence 180655,
41	39	69.6	247	10	US-09-975-719-363	Sequence 363, App
42	39	69.6	284	12	US-10-425-114-55247	Sequence 55247, A
43	39	69.6	391	16	US-10-437-963-142586	Sequence 142586,
44	39	69.6	447	16	US-10-437-963-166961	Sequence 166961,
45	39	69.6	674	10	US-09-975-719-51	Sequence 51, Appl

## ALIGNMENTS

RESULT 1  
US-10-615-959-42  
; Sequence 42, Application US/10615959  
; Publication NO. US20040048321A1  
; GENERAL INFORMATION:

; APPLICANT: Eyre, David R.  
; TITLE OF INVENTION: SYNTHETIC PEPTIDES FOR COLLAGEN RESORPTION ASSAYS  
; FILE REFERENCE: WRO-1-18220  
; CURRENT APPLICATION NUMBER: US/10/615,959  
; CURRENT FILING DATE: 2003-07-08  
; PRIOR APPLICATION NUMBER: US/10/009,999A  
; PRIOR FILING DATE: 2003-01-22  
; PRIOR APPLICATION NUMBER: PCT/US99/29357  
; PRIOR FILING DATE: 1999-12-10  
; PRIOR APPLICATION NUMBER: US 09/335,098  
; PRIOR FILING DATE: 1999-06-17  
; PRIOR APPLICATION NUMBER: US 60/141,574  
; PRIOR FILING DATE: 1999-06-29  
; PRIOR APPLICATION NUMBER: US 60/142,274  
; PRIOR FILING DATE: 1999-07-02  
; PRIOR APPLICATION NUMBER: US 60/142,675  
; PRIOR FILING DATE: 1999-07-07  
; PRIOR APPLICATION NUMBER: US 09/385,740  
; PRIOR FILING DATE: 1999-08-30  
; NUMBER OF SEQ ID NOS: 45  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 42  
; LENGTH: 11  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: MISC FEATURE  
; OTHER INFORMATION: synthetic peptide corresponding to carboxy-terminal  
; OTHER INFORMATION: telopeptide sequence of human type III collagen  
US-10-615-959-42

Query Match 100.0%; Score 56; DB 12; Length 11;  
Best Local Similarity 100.0%; Pred. No. 0.035;

Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 IAGIGGEKAGG 11  
|||||  
Db 1 IAGIGGEKAGG 11

## RESULT 2

US-10-615-959-43  
; Sequence 43, Application US/10615959  
; Publication No. US20040048321A1  
; GENERAL INFORMATION:  
; APPLICANT: Eyre, David R.  
; TITLE OF INVENTION: SYNTHETIC PEPTIDES FOR COLLAGEN RESORPTION ASSAYS  
; FILE REFERENCE: WPOS-1-18220  
; CURRENT APPLICATION NUMBER: US/10/615,959  
; CURRENT FILING DATE: 2003-07-08  
; PRIOR APPLICATION NUMBER: US/10/009,999A  
; PRIOR FILING DATE: 2003-01-22  
; PRIOR APPLICATION NUMBER: PCT/US99/29357  
; PRIOR FILING DATE: 1999-12-10  
; PRIOR APPLICATION NUMBER: US 09/335,098  
; PRIOR FILING DATE: 1999-06-17  
; PRIOR APPLICATION NUMBER: US 60/141,574  
; PRIOR FILING DATE: 1999-06-29  
; PRIOR APPLICATION NUMBER: US 60/142,274  
; PRIOR FILING DATE: 1999-07-02  
; PRIOR APPLICATION NUMBER: US 60/142,675  
; PRIOR FILING DATE: 1999-07-07  
; PRIOR APPLICATION NUMBER: US 09/385,740  
; PRIOR FILING DATE: 1999-08-30  
; NUMBER OF SEQ ID NOS: 45  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 43  
; LENGTH: 12  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; NAME/KEY: MISC FEATURE  
; OTHER INFORMATION: synthetic peptide corresponding to carboxy-terminal  
; OTHER INFORMATION: telopeptide sequence of human type III collagen  
US-10-615-959-43

Query Match 100.0%; Score 56; DB 12; Length 12;  
Best Local Similarity 100.0%; Pred. No. 0.038;

Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 IAGIGGEKAGG 11  
|||||  
Db 1 IAGIGGEKAGG 11

## RESULT 3

US-10-366-125-11  
; Sequence 11, Application US/10366125  
; Publication No. US20030228259A1  
; GENERAL INFORMATION:  
; APPLICANT: Hellerstein, Marc  
; TITLE OF INVENTION: MEASUREMENT OF BIOSYNTHESIS AND BREAKDOWN RATES OF  
; TITLE OF INVENTION: BIOLOGICAL MOLECULES THAT ARE INACCESSIBLE OR NOT  
; TITLE OF INVENTION: EASILY ACCESSIBLE TO DIRECT SAMPLING, NON-INVASIVELY,  
; TITLE OF INVENTION: BY LABEL INCORPORATION INTO METABOLIC DERIVATIVES AND  
; TITLE OF INVENTION: CATABOLITIC PRODUCTS  
; FILE REFERENCE: 416272003500  
; CURRENT APPLICATION NUMBER: US/10/366,125  
; CURRENT FILING DATE: 2003-02-12  
; PRIOR APPLICATION NUMBER: US 60/356,008  
; PRIOR FILING DATE: 2002-02-12  
; NUMBER OF SEQ ID NOS: 28  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 11  
; LENGTH: 25  
; TYPE: PRT

; ORGANISM: Homo sapiens  
US-10-366-125-11

Query Match 100.0%; Score 56; DB 15; Length 25;  
Best Local Similarity 100.0%; Pred. No. 0.079;

Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 IAGIGGEKAGG 11  
|||||  
Db 9 IAGIGGEKAGG 19

## RESULT 4

US-10-366-125-19  
; Sequence 19, Application US/10366125  
; Publication No. US20030228259A1  
; GENERAL INFORMATION:  
; APPLICANT: Hellerstein, Marc  
; TITLE OF INVENTION: MEASUREMENT OF BIOSYNTHESIS AND BREAKDOWN RATES OF  
; TITLE OF INVENTION: BIOLOGICAL MOLECULES THAT ARE INACCESSIBLE OR NOT  
; TITLE OF INVENTION: EASILY ACCESSIBLE TO DIRECT SAMPLING, NON-INVASIVELY,  
; TITLE OF INVENTION: BY LABEL INCORPORATION INTO METABOLIC DERIVATIVES AND  
; TITLE OF INVENTION: CATABOLITIC PRODUCTS  
; FILE REFERENCE: 416272003500  
; CURRENT APPLICATION NUMBER: US/10/366,125  
; CURRENT FILING DATE: 2003-02-12  
; PRIOR APPLICATION NUMBER: US 60/356,008  
; PRIOR FILING DATE: 2002-02-12  
; NUMBER OF SEQ ID NOS: 28  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 19  
; LENGTH: 262  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-366-125-19

Query Match 100.0%; Score 56; DB 15; Length 262;  
Best Local Similarity 100.0%; Pred. No. 0.82;

Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 IAGIGGEKAGG 11  
|||||  
Db 1 IAGIGGEKAGG 11

## RESULT 5

US-09-908-711-78  
; Sequence 78, Application US/09908711  
; Patent No. US20020045230A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: PA128  
; CURRENT APPLICATION NUMBER: US/09/908,711  
; CURRENT FILING DATE: 2001-07-20  
; PRIOR APPLICATION NUMBER: US01/01360  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: 09/764,867  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: US01/01344  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: 09/764,892  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: US01/01345  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: 09/764,888  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: US01/01329  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: 09/764,905  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: US01/01354  
; PRIOR FILING DATE: 2001-01-17

```

; PRIOR APPLICATION NUMBER: 09/764,891
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01339
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,869
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01340
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,874
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01334
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,898
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01320
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,853
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01349
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,902
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01239
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,870
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01348
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,882
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01347
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,896
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01307
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,864
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01341
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,856
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01336
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,868
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01312
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 60/179,065
; PRIOR FILING DATE: 2000-01-31
; PRIOR APPLICATION NUMBER: 60/180,628
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: 60/209,467
; PRIOR FILING DATE: 2000-06-07
; NUMBER OF SEQ ID NOS: 167
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 78
; LENGTH: 309
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (2)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (4)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-908-711-78

Query Match 100.0%; Score 56; DB 9; Length 309;
Best Local Similarity 100.0%; Pred. No. 0.97; Indels 0; Gaps 0;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

US-09-918-715-226
; Sequence 226, Application US/09918715
; Publication No. US20030017157A1
; GENERAL INFORMATION:
; APPLICANT: Brad St. Croix
; APPLICANT: Kenneth Kinzler
; TITLE OF INVENTION: ENDOTHELIAL CELL EXPRESSION PATTERNS
; FILE REFERENCE: 1107.00134
; CURRENT APPLICATION NUMBER: US/09/918,715
; CURRENT FILING DATE: 2001-08-01
; PRIOR APPLICATION NUMBER: 60/222,599
; PRIOR FILING DATE: 2000-08-02
; PRIOR APPLICATION NUMBER: 60/224,360
; PRIOR FILING DATE: 2000-08-11
; PRIOR APPLICATION NUMBER: 60/282,850
; PRIOR FILING DATE: 2000-04-11
; NUMBER OF SEQ ID NOS: 358
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 226
; LENGTH: 1466
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-918-715-226

Query Match 100.0%; Score 56; DB 12; Length 1466;
Best Local Similarity 100.0%; Pred. No. 4.6; Indels 0; Gaps 0;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IAGIGGEKAGG 11
DB 48 IAGIGGEKAGG 58

RESULT 6
US-09-918-715-226
; Sequence 226, Application US/09918715
; Publication No. US20030017157A1
; GENERAL INFORMATION:
; APPLICANT: Brad St. Croix
; APPLICANT: Kenneth Kinzler
; TITLE OF INVENTION: ENDOTHELIAL CELL EXPRESSION PATTERNS
; FILE REFERENCE: 1107.00134
; CURRENT APPLICATION NUMBER: US/09/918,715
; CURRENT FILING DATE: 2001-08-01
; PRIOR APPLICATION NUMBER: 60/222,599
; PRIOR FILING DATE: 2000-08-02
; PRIOR APPLICATION NUMBER: 60/224,360
; PRIOR FILING DATE: 2000-08-11
; PRIOR APPLICATION NUMBER: 60/282,850
; PRIOR FILING DATE: 2000-04-11
; NUMBER OF SEQ ID NOS: 358
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 226
; LENGTH: 1466
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-918-715-226

Query Match 100.0%; Score 56; DB 12; Length 1466;
Best Local Similarity 100.0%; Pred. No. 4.6; Indels 0; Gaps 0;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IAGIGGEKAGG 11
DB 1205 IAGIGGEKAGG 1215

RESULT 7
US-10-257-021-72
; Sequence 72, Application US/10257021
; Publication No. US20030211498A1
; GENERAL INFORMATION:
; APPLICANT: Morin, Patrice J.
; APPLICANT: Sherman-Baust, Cheryl A.
; APPLICANT: Pizer, Ellen S.
; APPLICANT: Hough, Colleen D.
; TITLE OF INVENTION: TUMOR MARKERS IN OVARIAN CANCER
; FILE REFERENCE: 14014.036902
; CURRENT APPLICATION NUMBER: US/10/257,021
; CURRENT FILING DATE: 2002-10-03
; PRIOR APPLICATION NUMBER: PCT/US01/10947
; PRIOR FILING DATE: 2001-04-03
; PRIOR APPLICATION NUMBER: 60/194,336
; PRIOR FILING DATE: 2000-04-03
; NUMBER OF SEQ ID NOS: 147
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 72
; LENGTH: 1466
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-257-021-72

Query Match 100.0%; Score 56; DB 12; Length 1466;
Best Local Similarity 100.0%; Pred. No. 4.6; Indels 0; Gaps 0;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IAGIGGEKAGG 11
DB 1205 IAGIGGEKAGG 1215
```

```

; TITLE OF INVENTION: COMPOSITIONS, KITS, AND METHODS FOR IDENTIFICATION, ASSESSMENT,
; FILE REFERENCE: MRI-038
; CURRENT FILING DATE: 2002-06-21
; PRIOR APPLICATION NUMBER: US 60/299,887
; PRIOR FILING DATE: 2001-06-21
; PRIOR APPLICATION NUMBER: US 60/301,572
; PRIOR FILING DATE: 2001-06-27
; PRIOR APPLICATION NUMBER: US 60/306,501
; PRIOR FILING DATE: 2001-07-18
; PRIOR APPLICATION NUMBER: US 60/325,002
; PRIOR FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: US 60/362,585
; PRIOR FILING DATE: 2002-03-05
; PRIOR APPLICATION NUMBER: US 60/xxx,xxx
; PRIOR FILING DATE: 2002-05-14
; NUMBER OF SEQ ID NOS: 506
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 68
; LENGTH: 1466
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-177-293-68

Query Match      100.0%; Score 56; DB 14; Length 1466;
Best Local Similarity 100.0%; Pred. No. 4.6;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 IAGIGGEKAGG 11
Db      1205 IAGIGGEKAGG 1215

RESULT 8
US-10-177-293-68
; Sequence 68, Application US/10177293
; Publication No. US20030124128A1
; GENERAL INFORMATION:
; APPLICANT: Lillie, James
; APPLICANT: Glatt, Karen
; APPLICANT: Zhao, Xumei
; APPLICANT: Gannavaru, Manjula
; APPLICANT: Kamatkar, Shubhangi
; APPLICANT: Mertens, Maureen
; APPLICANT: Myer, Vic
; APPLICANT: Wang, Youzhen
; APPLICANT: Xu, Yongyao
; APPLICANT: Hoersch, Sebastian
; APPLICANT: Monahan, John
; APPLICANT: Meyers, Rachel E.
; APPLICANT: Bast Jr., Robert C.
; APPLICANT: Hortobagyi, Gabriel N.
; APPLICANT: Pusztai, Lajos
; APPLICANT: Meric, Funda
; APPLICANT: Sabin, Aysegul
; APPLICANT: Mills, Gordon B.
; TITLE OF INVENTION: COMPOSITIONS, KITS, AND METHODS FOR IDENTIFICATION, ASSESSMENT,
; FILE REFERENCE: MRI-038
; CURRENT FILING DATE: 2002-06-21
; PRIOR APPLICATION NUMBER: US 60/299,887
; PRIOR FILING DATE: 2001-06-21
; PRIOR APPLICATION NUMBER: US 60/301,572
; PRIOR FILING DATE: 2001-06-27
; PRIOR APPLICATION NUMBER: US 60/306,501
; PRIOR FILING DATE: 2001-07-18
; PRIOR APPLICATION NUMBER: US 60/325,002
; PRIOR FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: US 60/362,585
; PRIOR FILING DATE: 2002-03-05
; PRIOR APPLICATION NUMBER: US 60/xxx,xxx
; PRIOR FILING DATE: 2002-05-14
; NUMBER OF SEQ ID NOS: 506
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 68
; LENGTH: 1466
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-177-293-68

```

```

Query Match      100.0%; Score 56; DB 14; Length 1466;
Best Local Similarity 100.0%; Pred. No. 4.6;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 IAGIGGEKAGG 11
Db      1205 IAGIGGEKAGG 1215

```

```

RESULT 9
US-10-301-822-33
; Sequence 33, Application US/10301822
; Publication No. US20030148410A1
; GENERAL INFORMATION:
; APPLICANT: Millennium Pharmaceuticals, Inc.
; APPLICANT: Berger, Allison
; APPLICANT: Guillemette, Tracy L.
; APPLICANT: Kamatkar, Shubhangi
; APPLICANT: Schlegel, Robert
; APPLICANT: Monahan, John E.
; APPLICANT: Thibodeau, Stephen N.
; APPLICANT: Burgart, Lawrence J.
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND
; FILE REFERENCE: METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND

```

```

; TITLE OF INVENTION: THERAPY OF COLON CANCER
; FILE REFERENCE: MPM01-029P2RNM
; CURRENT APPLICATION NUMBER: US/10/301,822
; CURRENT FILING DATE: 2002-11-21
; PRIOR APPLICATION NUMBER: US 60/339,971
; PRIOR FILING DATE: 2001-12-10
; PRIOR APPLICATION NUMBER: US 60/361,978
; PRIOR FILING DATE: 2002-03-05
; PRIOR APPLICATION NUMBER: US 60/381,988
; PRIOR FILING DATE: 2002-05-20
; NUMBER OF SEQ ID NOS: 228
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 33
; LENGTH: 1466
; TYPE: PRT
; ORGANISM: Homo Sapiens
US-10-301-822-33

Query Match      100.0%; Score 56; DB 14; Length 1466;
Best Local Similarity 100.0%; Pred. No. 4.6;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 IAGIGGEKAGG 11
Db      1205 IAGIGGEKAGG 1215

RESULT 10
US-10-357-851-3
; Sequence 3, Application US/10357851
; Publication No. US20040151731A1
; GENERAL INFORMATION:
; APPLICANT: Jicha, Douglas L.
; TITLE OF INVENTION: Method and Compositions Involving
; FILE REFERENCE: T-Lymphocyte Reactivity with Collagen in Aortic Tissue of
; TITLE OF INVENTION: Abdominal Aortic Aneurysm Patients
; FILE REFERENCE: 13376US
; CURRENT APPLICATION NUMBER: US/10/357,851
; CURRENT FILING DATE: 2003-02-04
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 1466
; TYPE: PRT
; ORGANISM: Homo Sapiens
US-10-357-851-3

Query Match      100.0%; Score 56; DB 16; Length 1466;
Best Local Similarity 100.0%; Pred. No. 4.6;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 IAGIGGEKAGG 11
Db      1205 IAGIGGEKAGG 1215

RESULT 11
US-10-358-024-3
; Sequence 3, Application US/10358024
; Publication No. US20040151732A1
; GENERAL INFORMATION:
; APPLICANT: Jicha, Douglas L.
; APPLICANT: Felsue, Stephen
; TITLE OF INVENTION: Methods and Compositions Involving Blood
; FILE REFERENCE: T-Lymphocytes Reactivity with Collagen
; FILE REFERENCE: 13436US
; CURRENT APPLICATION NUMBER: US/10/358,024
; CURRENT FILING DATE: 2003-02-04
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 1466
; TYPE: PRT

```

```
; ORGANISM: Homo Sapiens
US-10-358-024-3
Query Match      100.0%; Score 56; DB 16; Length 1466;
Best Local Similarity 100.0%; Pred. No. 4.6;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 IAGIGGEKAGG 11
Db      1205 IAGVGGEKAGG 1215

RESULT 12
US-10-734-564-103
; Sequence 103, Application US/10734564
; Publication No. US20040157278A1
; GENERAL INFORMATION:
; APPLICANT: Christopher C Burgess et al
; TITLE OF INVENTION: Detection Methods Using TIMP1
; FILE REFERENCE: 1657/2012
; CURRENT APPLICATION NUMBER: US/10/734,564
; CURRENT FILING DATE: 2003-12-12
; NUMBER OF SEQ ID NOS: 138
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 103
; LENGTH: 1466
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-734-564-103

Query Match      100.0%; Score 56; DB 16; Length 1466;
Best Local Similarity 100.0%; Pred. No. 4.6;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 IAGIGGEKAGG 11
Db      1205 IAGVGGEKAGG 1215

RESULT 13
US-10-402-089-12
; Sequence 12, Application US/10402089
; Publication No. US20040005663A1
; GENERAL INFORMATION:
; APPLICANT: Bell, Marcum P.
; APPLICANT: Neff, Thomas B.
; APPLICANT: Polarek, James W.
; APPLICANT: Seeley, Todd W.
; TITLE OF INVENTION: PORCINE COLLAGENS AND GELATINS
; FILE REFERENCE: FP0402.3 CON
; CURRENT APPLICATION NUMBER: US/10/402,089
; CURRENT FILING DATE: 2003-03-26
; PRIOR APPLICATION NUMBER: US 09/709,700
; PRIOR FILING DATE: 2000-11-10
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 12
; LENGTH: 1466
; TYPE: PRT
; ORGANISM: Sus scrofa
US-10-402-089-12

Query Match      98.2%; Score 55; DB 15; Length 1466;
Best Local Similarity 90.9%; Pred. No. 6.5;
Matches 10; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      1 IAGIGGEKAGG 11
Db      1205 IAGVGGEKAGG 1215

RESULT 14
US-10-402-072A-12
; Sequence 12, Application US/10402072A
; Publication No. US20040018592A1
; GENERAL INFORMATION:
; APPLICANT: Bell, Marcum P.
; APPLICANT: Neff, Thomas B.
; APPLICANT: Polarek, James W.
; APPLICANT: Seeley, Todd W.
; TITLE OF INVENTION: BOVINE COLLAGENS AND GELATINS
; FILE REFERENCE: FP0402.2 CON
; CURRENT APPLICATION NUMBER: US/10/402,072A
; CURRENT FILING DATE: 2003-03-26
; PRIOR APPLICATION NUMBER: US 09/709,700
; PRIOR FILING DATE: 2000-11-10
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 12
; LENGTH: 1466
; TYPE: PRT
; ORGANISM: Sus scrofa
US-10-402-072A-12

Query Match      98.2%; Score 55; DB 15; Length 1466;
Best Local Similarity 90.9%; Pred. No. 6.5;
Matches 10; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      1 IAGIGGEKAGG 11
Db      1205 IAGVGGEKAGG 1215

RESULT 15
US-10-058-124-21
; Sequence 21, Application US/10058124
; Publication No. US20030119058A1
; GENERAL INFORMATION:
; APPLICANT: Qvist, Per
; APPLICANT: Bonde, Martin
; TITLE OF INVENTION: A Method for Assaying Collagen Fragments
; in Body Fluids, A Test Kit and Means for Carrying Out the
; Method and Use of the Method to Diagnose the Presence of
; Disorders Associated with the Metabolism of

NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSER: Darby & Darby PC
STREET: 805 Third Avenue
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10022

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/058,124
FILING DATE: 29-Jan-2002
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/570,573
FILING DATE: 2002-MAY-12
APPLICATION NUMBER: 08/187,319
FILING DATE: <Unknown>

ATTORNEY/AGENT INFORMATION:
NAME: Gogoris, Adda C
REGISTRATION NUMBER: 29,714
REFERENCE/DOCKET NUMBER: 4305/08701
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-527-7700
TELEFAX: 212-753-6237
TELEX: 236687

INFORMATION FOR SEQ ID NO: 21:
SEQUENCE CHARACTERISTICS:
```

```

; LENGTH: 1078 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; IMMEDIATE SOURCE:
; CLONE: COLLAGEN ALPHA 1 (III)
; SEQUENCE DESCRIPTION: SEQ ID NO: 21:
US-10-058-124-21

```

```

Query Match      89.3%; Score 50; DB 14; Length 1078;
Best Local Similarity 90.9%; Pred. No. 28;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

QY      1 IAGIGGKAGG 11
      |||||
Db      1058 IAGIGAERKAGG 1068

```

Search completed: September 18, 2004, 04:56:13  
Job time : 52.1961 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 18, 2004, 04:02:05 ; Search time 13.3725 Seconds  
(without alignments)  
79.125 Million cell updates/sec

Title: US-10-615-959-42  
Perfect score: 56  
Sequence: 1 IAGIGGEKAGG 11

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR 78:\*  
1: piri:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	56	100.0	1466	1 CGHU7L	collagen alpha 1(I
2	52	92.9	636	2 S41067	collagen alpha 1(I
3	52	92.9	1464	2 S59856	collagen alpha 1(I
4	44	78.6	161	2 S14998	heat shock protein
5	44	78.6	487	2 F70863	probable folypoly
6	42	75.0	209	2 AE2981	arylesterase (impo
7	42	75.0	209	2 C98302	PBSX prophage ORF
8	41	73.2	1332	2 F69732	formylmethanofuran
9	40	71.4	270	2 H69074	formylmethanofuran
10	40	71.4	270	2 S57457	hypothetical prote
11	40	71.4	298	2 E95286	hypothetical prote
12	40	71.4	548	2 T51035	hypothetical prote
13	39	69.6	87	2 T14302	glycine-rich cell
14	39	69.6	163	2 T47394	hypothetical prote
15	39	69.6	399	2 I49754	homeobox protein -
16	39	69.6	846	2 H70599	hypothetical prote
17	38	67.9	151	2 T08002	glycine-rich prote
18	38	67.9	205	2 F84274	halocyanin precurs
19	38	67.9	378	1 OZZQAL	circumsporozoite p
20	38	67.9	423	2 B88450	protein F21H11.3 [
21	38	67.9	495	2 D83167	leucine aminopepti
22	38	67.9	588	2 F70971	hypothetical glyci
23	38	67.9	605	2 H95240	conserved hypothec
24	38	67.9	605	2 A99705	hypothetical prote
25	38	67.9	694	2 T00148	hypothetical prote
26	38	67.9	875	2 T30023	hypothetical prote
27	38	67.9	935	2 T29390	hypothetical prote
28	38	67.9	1545	2 T14366	phage lambda-relat
29	37	66.1	126	2 A11951	hypothetical prote

coat protein - pas  
probable ATP-bind  
conserved hypothec  
tRNA (5-methylamin  
anthranilate synth  
hypothetical prote  
dihydrolipoamide d  
NF2d9 - mouse  
metalloproteinase  
neutral proteinase  
hypothetical prote  
protein M28.1 limp  
probable homeodoma  
probable protein k  
low-density lipopr  
furin (EC 3.4.21.7

## ALIGNMENTS

### RESULT 1

CGHU7L

collagen alpha 1(III) chain precursor - human

N/Alternate names: procollagen alpha 1(III) chain

C/Species: Homo sapiens (man)

C/Date: 24-Apr-1984 #sequence revision 01-Sep-1995 #text change 21-Jul-2000

C/Accession: S05272; S04642; PE0011; S01726; S04887; A90399; A94562; I51868; S59511; A904

R/Prockop, D.J.

submitted to the EMBL Data Library, February 1989

A/Reference number: S05272

A/Accession: S05272

A/Status: preliminary

A/Molecule type: mRNA

A/Residues: 1-1240,'V',1242-1466 <PRC>

A/Cross-references: EMBL:X14420; NID:G30057; PIDN:CAA32583.1; PID:G30058

R/Ala-Kokko, L.; Kontusaari, S.; Baldwin, C.T.; Kuivaniemi, H.; Prockop, D.J.

Biochem. J. 260, 509-516, 1989

A/Title: Structure of cDNA clones coding for the entire prepro-alpha(III) chain of human  
erences.

A/Reference number: S04642; MUID:89350838; PMID:2764886

A/Accession: S04642

A/Molecule type: mRNA

A/Residues: 1-1196 <ALA>

A/Cross-references: EMBL:X14420; NID:G30057; PIDN:CAA32583.1; PID:G30058

A/Note: the complete sequence is not shown

R/Benson-Chanda, V.; Su, M.W.; Weil, D.; Chu, M.L.; Ramirez, F.

Gene 78, 255-265, 1989

A/Title: Cloning and analysis of the 5' portion of the human type-III procollagen gene (C

A/Reference number: PE0011; MUID:89378752; PMID:2777083

A/Accession: PE0011

A/Molecule type: DNA

A/Residues: 1-176 <BEN>

A/Cross-references: GB:M26939; NID:G180813; PIDN:AAA52040.1; PID:G180814

R/Toman, P.D.; Ricca, G.A.; de Crombrughe, B.

Nucleic Acids Res. 16, 7201, 1988

A/Title: Nucleotide sequence of a cDNA coding for the amino-terminal region of human pref

A/Reference number: S01726; MUID:88303360; PMID:3405773

A/Accession: S01726

A/Molecule type: mRNA

A/Residues: 1-170 <TOM>

A/Cross-references: EMBL:X07240; NID:G30060; PIDN:CAA30229.1; PID:G30061

A/Note: the authors translated the codon CAG for residue 154 as His

R/Janczko, R.A.; Ramirez, F.

Nucleic Acids Res. 17, 6742, 1989

A/Title: Nucleotide and amino acid sequences of the entire human alpha-1 (III) collagen.

A/Reference number: S04887; MUID:89386015; PMID:2780304

A/Accession: S04887

A/Molecule type: mRNA

A/Residues: 149-163,'G',164-240,'D',242-471,'D',473-487,'L',489,'S',491-613,'Y',615-634,'

A/Cross-references: EMBL:X15332; NID:G29545; PIDN:CAA33387.1; PID:G930045

A/Note: the authors' translation of residues 905-932 is inconsistent with the nucleotide

R/Seyer, J.M.; Kang, A.H.

Biochemistry 16, 1158-1164, 1977  
A:Title: Covalent structure of collagen: amino acid sequence of cyanogen bromide peptide  
A:Reference number: A90399; MUID:77134724; PMID:557335  
A:Accession: A90399  
A:Molecule type: protein  
A:Residues: 'V', 159-225, 229-232, 'P', 234-292, 'D', 294-398 <SEY1>  
A:Experimental source: liver  
A:Note: sequence corrected by A94562; attachment of 2-O-alpha-D-glucosyl-O-beta-D-galactose submitted to the Atlas, December 1977  
R:Seyer, J.M.  
A:Reference number: A94562  
A:Accession: A94562  
A:Molecule type: protein  
A:Residues: 'V', 169-225, 229-277, 'A', 279-292, 'D', 294, 'S', 296-398 <SEY2>  
A:Experimental source: liver  
A:Note: author submitted corrections to A90399  
R:Miliewicz, D.M.; Witz, A.M.; Smith, A.C.; Manchester, D.K.; Waldstein, G.; Byers, P.H. Am. J. Hum. Genet. 53, 62-70, 1993  
A:Title: Parental somatic and germ-line mosaicism for a multiexon deletion with unusual splicing.  
A:Reference number: I51868; MUID:93304430; PMID:8317500  
A:Accession: I51868  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 186-194 <MTL>  
A:Cross-references: GB:S62925; NID:g386425; PIDN:AAD13937.1; PID:g4261637  
R:Chiodo, A.A.; Silience, D.O.; Cole, W.G.; Bateman, J.F. Biochem. J. 311, 939-943, 1995  
A:Title: Abnormal type III collagen produced by an exon-17-skipping mutation of the COL3A1 gene.  
A:Reference number: S59511; MUID:96067614; PMID:7487954  
A:Accession: S59511  
A:Molecule type: mRNA  
A:Residues: 302-423 <CHI>  
A:Cross-references: GB:S79877; NID:gl195576; PIDN:AAB35615.1; PID:gl195577  
R:Seyer, J.M.; Kang, A.H. Biochemistry 17, 3404-3411, 1978  
A:Title: Covalent structure of collagen: amino acid sequence of five consecutive CNBr peptides.  
A:Reference number: A90414; MUID:79000343; PMID:687591  
A:Accession: A90414  
A:Molecule type: protein  
A:Residues: 399-675, 'N', 677-727 <SEY3>  
A:Experimental source: liver  
R:Lee, B.; Vitale, E.; Superiti-Furga, A.; Steinmann, B.; Ramirez, F. J. Biol. Chem. 266, 5256-5259, 1991  
A:Title: G to T transversion at position +5 of a splice donor site causes skipping of the exon 19.  
A:Reference number: I55349; MUID:91161621; PMID:1672129  
A:Accession: I55349  
A:Status: translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 537-605 <LEE>  
A:Cross-references: GB:M59312; NID:gl180815; PIDN:AAA52041.1; PID:gl180816  
R:Seyer, J.M.; Mainardi, C.; Kang, A.H. Biochemistry 19, 1583-1589, 1980  
A:Title: Covalent structure of collagen: amino acid sequence of alpha1 (III)-CB5 from tyrosine.  
A:Reference number: A90438; MUID:80198282; PMID:6246925  
A:Accession: A90438  
A:Molecule type: protein  
A:Residues: 728-895, 'A', 897-964 <SEY4>  
A:Experimental source: liver  
R:Cole, W.G.; Chiodo, A.A.; Lamande, S.R.; Janeczko, R.; Ramirez, F.; Dahl, H.H.M.; Chan J. Biol. Chem. 265, 17070-17077, 1990  
A:Title: A base substitution at a splice site in the COL3A1 gene causes exon skipping and exon skipping.  
A:Reference number: A38303; MUID:91009133; PMID:2145268  
A:Accession: A38303  
A:Molecule type: mRNA  
A:Residues: 861-1015 <COL>  
A:Cross-references: GB:J05617; GB:M55603; GB:M59227; NID:gl180878; PIDN:AAB59393.1; PID:g180878  
A:Note: a mutant sequence with 942-977 spliced out from a patient with Ehlers-Danlos syndrome.  
R:Mankoo, B.S.; Dalgleish, R. Nucleic Acids Res. 16, 2337, 1988  
A:Title: Human pro alpha1(III) collagen: cDNA sequence for the 3' end.  
A:Reference number: S02119; MUID:88189827; PMID:3357782  
A:Accession: S02119

A>Status: translation not shown

A:Molecule type: mRNA

A:Residues: 950-1018,'Y',1020-1183,'S','S',1185-1466 <MAN>

A:Cross-references: EMBL:X06700; NID:G30053; PIDN:CAA29886.1; PID:G30054

R:Seyer, J.M.; Kang, A.H.

Biochemistry 20, 2621-2627, 1981

A>Title: Covalent structure of collagen: amino acid sequence of alpha (III)-CB9 from tnyt

A:Reference number: A90446; MUID:81208139; PMID:7016180

A:Accession: A90446

A:Molecule type: protein

A:Residues: 965-979,'A',981-984,'PS',987,'QN',990-1096,'P',1098-1152,'AT',1155,'S',1157-1159

A:Experimental source: liver

R:Loidl, H.R.; Brinker, J.M.; May, M.; Pihlajanien, T.; Morrow, S.; Rosenbloom, J.; Mye

Nucleic Acids Res. 12, 9383-9394, 1984

A>Title: Molecular cloning and carboxyl-propeptide analysis of human type III procollage

A:Reference number: A93551; MUID:85087944; PMID:6096827

A:Accession: A93551

A:Molecule type: mRNA

A:Residues: 1065-1155,'P',1157-1466 <LOI>

A:Cross-references: EMBL:X01655; EMBL:X01742; NID:G29584; PIDN:CAA25821.1

R:Mirskulin, M.; Dalgleish, R.; Kluge-Beckerman, B.; Rennard, S.I.; Tolstoshev, P.; Brant

Biochemistry 25, 1408-1413, 1986

A>Title: Human type III collagen gene expression is coordinately modulated with the type

A:Reference number: I52393; MUID:86187804; PMID:3754462

A:Accession: I52393

A>Status: translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1161-1200 <MS>

A:Cross-references: GB:M13146; NID:G180415; PIDN:AAAS2003.1; PID:g180416

R:Emanuel, B.S.; Cannizzaro, L.A.; Seyer, J.M.; Myers, J.C.

Proc. Natl. Acad. Sci. U.S.A. 82, 3385-3389, 1985

A>Title: Human alpha 1(III) and alpha 2(V) procollagen genes are located on the long arm

A:Reference number: I59025; MUID:85216505; PMID:3858826

A:Accession: I79359

A>Status: translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1165-1196 <EMA>

A:Cross-references: GB:M11134; NID:G180417; PIDN:AAAS2004.1; PID:g180418

R:Chu, M.L.; Weil, D.; de Wet, W.; Bernard, M.; Sippola, M.; Ramirez, F.

J. Biol. Chem. 260, 4357-4363, 1985

A>Title: Isolation of cDNA and genomic clones encoding human pro-alpha(III) collagen. P

A:Reference number: A92516; MUID:85157600; PMID:2579949

A:Accession: A92516

A:Molecule type: DNA

A:Residues: 1176-1240,'V',1242-1356,'P',1358-1466 <CHU>

A:Cross-references: GB:M10615; GB:M10793; GB:M10794; GB:M10795; GB:M10796; GB:M10797; GB

A:Experimental source: liver

A>Note: the authors translated the codon TTC for residue 1057 as Tyr; the codons given for

ation

C:Comment: Prolines and lysines at the third position of the tripeptide repeating unit (G

3-hydroxylated. About 15% of the lysines are 5-hydroxylated and some are subsequently O

C:Genetics:

A:Gene: GDB:COL3A1

A:Cross-references: GDB:118729; OMIM:120180

A:Map position: 2q31-2q31

A:Introns: 27/1; 94/3; 111/3; 149/3; 176/3; 554/3; 587/3; 1175/3; 1275/1; 1337/3; 1418/3

A>Note: the list of introns is incomplete; defects in this gene can result in Ehlers-Dan

C:Complex: type III collagen is a homotrimer of monomers initially linked by disulfide b

er of their length, is formed with desmosine cross-links made from lysine and allysine re

C:Function:

A:Description: structural component of extracellular fibrous polymer that maintains integ

C:Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology; i

C:Keywords: coiled coil; Ehlers-Danlos syndrome; extracellular matrix; glycoprotein; hyd

F:1-23/Domain: signal sequence #status predicted <SIG>

F:24-153/Domain: amino-terminal propeptide #status predicted <PRO>

F:31-91/Domain: von Willebrand factor type C repeat homology <VWC>

F:154-1221/Product: collagen alpha 1(III) chain #status predicted <MAT>

F:154-167/Region: amino-terminal nonhelical telopeptide

F:168-1196/Region: helical

F:1091-1093/Region: cell attachment (R-G-D) motif

F:1197-1221/Region: carboxyl-terminal nonhelical telopeptide

F:1222-1466/Domain: carboxyl-terminal propeptide #status predicted <CPR>

F:1238-1466/Domain: fibrillar collagen carboxyl-terminal homology <CCR>

F:153/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status predicted  
 F:154/Cleavage site: Pro-Gln (procollagen N-endopeptidase) #status predicted  
 F:154/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status predicted  
 F:161/1212/Modified site: allysine (Lys) #status predicted  
 F:263/284,860,977,1106/Modified site: 5-hydroxylysine (Lys) #status experimental  
 F:263/Binding site: carbohydrate (Lys) (covalent) #status experimental  
 F:584,1094/Modified site: 5-hydroxylysine (Lys) (partial) #status experimental  
 F:948-949/Cleavage site: Gly-Ile (collagenase) #status experimental  
 F:1106/Binding site: carbohydrate (Lys) (covalent) #status predicted

Query Match 100.0%; Score 56; DB 1; Length 1466;  
 Best Local Similarity 100.0%; Pred. No. 0.35; Indels 0; Gaps 0;  
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IAGIGGEKAGG 11  
 |||:|||||  
 Db 1205 IAGVGGEKSGG 1215

RESULT 2  
 S41067  
 collagen alpha 1(III) chain - rat  
 C:Species: Rattus norvegicus (Norway rat)  
 C:Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 13-Aug-1999  
 C:Accession: S41067; A29905; S31924  
 R:Glumoff, V.; Maekelae, J.K.; Vuorio, E.  
 Biochim. Biophys. Acta 1217, 41-48, 1994  
 A:Title: Cloning of cDNA for rat pro alpha-1(III) collagen mRNA. Different expression pa  
 A:Reference number: S41067; MUID:94114571; PMID:8286415  
 A:Accession: S41067  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-636 <GLU>  
 A:Cross-references: EMBL:X70369; NID:G57915; PIDN:CAA49032.1; PID:G57916  
 R:Frankel, F.R.; Hsu, C.Y.J.; Myers, J.C.; Lin, E.; Lyttle, C.R.; Komm, B.; Mohn, K.  
 DNA 7, 347-354, 1988  
 A:Title: Regulation of alpha-2 (I), alpha-1 (III), and alpha-2(V) collagen mRNAs by est  
 A:Reference number: A29905; MUID:88296083; PMID:2456904  
 A:Accession: A29905  
 A:Molecule type: mRNA  
 A:Residues: 308-482 <PRA>  
 A:Cross-references: GB:M21354; NID:G203500; PIDN:AAA40942.1; PID:G203501  
 R:Glumoff, V.; Maekelae, J.K.; Vuorio, E.  
 submitted to the EMBL Data Library, February 1993  
 A:Reference number: S31924  
 A:Accession: S31924  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 2-636 <GL2>  
 A:Cross-references: EMBL:X70369  
 C:Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;  
 C:Keywords: coiled coil; extracellular matrix; glycoprotein; trimer; triple helix  
 F:408-636/Domain: fibrillar collagen carboxyl-terminal homology <FCC>

Query Match 92.9%; Score 52; DB 2; Length 636;  
 Best Local Similarity 81.8%; Pred. No. 0.73; Indels 0; Gaps 0;  
 Matches 9; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 IAGIGGEKAGG 11  
 |||:|||||  
 Db 375 IAGVGGEKSGG 385

RESULT 3  
 S59856  
 collagen alpha 1(III) chain precursor - mouse  
 C:Species: Mus musculus (house mouse)  
 C:Date: 10-Apr-1996 #sequence\_revision 19-Apr-1996 #text\_change 13-Aug-1999  
 C:Accession: S59856; S62120; S16373  
 R:Toman, P.D.; de Crombrughe, B.  
 Gene 147, 161-168, 1994  
 A:Title: The mouse type-III procollagen-encoding gene: genomic cloning and complete DNA  
 A:Reference number: S59856; MUID:95011609; PMID:7926795

A:Accession: S59856  
 A:Molecule type: DNA  
 A:Residues: 1-1464 <TOM>  
 A:Cross-references: EMBL:X52046  
 R:Toman, D.  
 submitted to the EMBL Data Library, November 1994  
 A:Reference number: S62120  
 A:Accession: S62120  
 A:Molecule type: DNA  
 A:Residues: 1-866, 'G', 868-1464 <TOA>  
 A:Cross-references: EMBL:X52046; NID:G575321; PIDN:CAA36279.1; PID:G575322  
 R:Metzger-Rantcho, M.; Toman, D.; de Crombrughe, B.; Vuorio, E.  
 Biochim. Biophys. Acta 1089, 241-243, 1991  
 A:Title: Specific hybridization probes for mouse type I, II, III and IX collagen mRNAs.  
 A:Reference number: S16176; MUID:91274355; PMID:2054384  
 A:Accession: S16373  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1442-1464 <MET>  
 A:Cross-references: EMBL:X57983; NID:G50476; PIDN:CAA41048.1; PID:G50477  
 C:Genetics:  
 A:Introns: 29/1; 95/3; 112/3; 150/3; 175/3; 193/3; 211/3; 229/3; 247/3; 265/3; 283/3; 298/3; 316/3; 334/3; 352/3; 370/3; 388/3; 406/3; 424/3; 442/3; 460/3; 478/3; 496/3; 514/3; 532/3; 550/3; 568/3; 586/3; 604/3; 622/3; 640/3; 658/3; 676/3; 694/3; 712/3; 730/3; 748/3; 766/3; 784/3; 802/3; 820/3; 838/3; 856/3; 874/3; 892/3; 910/3; 928/3; 946/3; 964/3; 982/3; 1000/3; 1018/3; 1036/3; 1054/3; 1072/3; 1090/3; 1108/3; 1126/3; 1144/3; 1162/3; 1180/3; 1198/3; 1216/3; 1234/3; 1252/3; 1270/3; 1288/3; 1306/3; 1324/3; 1342/3; 1360/3; 1378/3; 1396/3; 1414/3; 1432/3; 1450/3; 1468/3; 1486/3; 1504/3; 1522/3; 1540/3; 1558/3; 1576/3; 1594/3; 1612/3; 1630/3; 1648/3; 1666/3; 1684/3; 1702/3; 1720/3; 1738/3; 1756/3; 1774/3; 1792/3; 1810/3; 1828/3; 1846/3; 1864/3; 1882/3; 1900/3; 1918/3; 1936/3; 1954/3; 1972/3; 1990/3; 2008/3; 2026/3; 2044/3; 2062/3; 2080/3; 2098/3; 2116/3; 2134/3; 2152/3; 2170/3; 2188/3; 2206/3; 2224/3; 2242/3; 2260/3; 2278/3; 2296/3; 2314/3; 2332/3; 2350/3; 2368/3; 2386/3; 2404/3; 2422/3; 2440/3; 2458/3; 2476/3; 2494/3; 2512/3; 2530/3; 2548/3; 2566/3; 2584/3; 2602/3; 2620/3; 2638/3; 2656/3; 2674/3; 2692/3; 2710/3; 2728/3; 2746/3; 2764/3; 2782/3; 2800/3; 2818/3; 2836/3; 2854/3; 2872/3; 2890/3; 2908/3; 2926/3; 2944/3; 2962/3; 2980/3; 2998/3; 3016/3; 3034/3; 3052/3; 3070/3; 3088/3; 3106/3; 3124/3; 3142/3; 3160/3; 3178/3; 3196/3; 3214/3; 3232/3; 3250/3; 3268/3; 3286/3; 3304/3; 3322/3; 3340/3; 3358/3; 3376/3; 3394/3; 3412/3; 3430/3; 3448/3; 3466/3; 3484/3; 3502/3; 3520/3; 3538/3; 3556/3; 3574/3; 3592/3; 3610/3; 3628/3; 3646/3; 3664/3; 3682/3; 3700/3; 3718/3; 3736/3; 3754/3; 3772/3; 3790/3; 3808/3; 3826/3; 3844/3; 3862/3; 3880/3; 3898/3; 3916/3; 3934/3; 3952/3; 3970/3; 3988/3; 4006/3; 4024/3; 4042/3; 4060/3; 4078/3; 4096/3; 4114/3; 4132/3; 4150/3; 4168/3; 4186/3; 4204/3; 4222/3; 4240/3; 4258/3; 4276/3; 4294/3; 4312/3; 4330/3; 4348/3; 4366/3; 4384/3; 4402/3; 4420/3; 4438/3; 4456/3; 4474/3; 4492/3; 4510/3; 4528/3; 4546/3; 4564/3; 4582/3; 4600/3; 4618/3; 4636/3; 4654/3; 4672/3; 4690/3; 4708/3; 4726/3; 4744/3; 4762/3; 4780/3; 4798/3; 4816/3; 4834/3; 4852/3; 4870/3; 4888/3; 4906/3; 4924/3; 4942/3; 4960/3; 4978/3; 4996/3; 5014/3; 5032/3; 5050/3; 5068/3; 5086/3; 5104/3; 5122/3; 5140/3; 5158/3; 5176/3; 5194/3; 5212/3; 5230/3; 5248/3; 5266/3; 5284/3; 5302/3; 5320/3; 5338/3; 5356/3; 5374/3; 5392/3; 5410/3; 5428/3; 5446/3; 5464/3; 5482/3; 5500/3; 5518/3; 5536/3; 5554/3; 5572/3; 5590/3; 5608/3; 5626/3; 5644/3; 5662/3; 5680/3; 5698/3; 5716/3; 5734/3; 5752/3; 5770/3; 5788/3; 5806/3; 5824/3; 5842/3; 5860/3; 5878/3; 5896/3; 5914/3; 5932/3; 5950/3; 5968/3; 5986/3; 6004/3; 6022/3; 6040/3; 6058/3; 6076/3; 6094/3; 6112/3; 6130/3; 6148/3; 6166/3; 6184/3; 6202/3; 6220/3; 6238/3; 6256/3; 6274/3; 6292/3; 6310/3; 6328/3; 6346/3; 6364/3; 6382/3; 6400/3; 6418/3; 6436/3; 6454/3; 6472/3; 6490/3; 6508/3; 6526/3; 6544/3; 6562/3; 6580/3; 6598/3; 6616/3; 6634/3; 6652/3; 6670/3; 6688/3; 6706/3; 6724/3; 6742/3; 6760/3; 6778/3; 6796/3; 6814/3; 6832/3; 6850/3; 6868/3; 6886/3; 6904/3; 6922/3; 6940/3; 6958/3; 6976/3; 6994/3; 7012/3; 7030/3; 7048/3; 7066/3; 7084/3; 7102/3; 7120/3; 7138/3; 7156/3; 7174/3; 7192/3; 7210/3; 7228/3; 7246/3; 7264/3; 7282/3; 7300/3; 7318/3; 7336/3; 7354/3; 7372/3; 7390/3; 7408/3; 7426/3; 7444/3; 7462/3; 7480/3; 7498/3; 7516/3; 7534/3; 7552/3; 7570/3; 7588/3; 7606/3; 7624/3; 7642/3; 7660/3; 7678/3; 7696/3; 7714/3; 7732/3; 7750/3; 7768/3; 7786/3; 7804/3; 7822/3; 7840/3; 7858/3; 7876/3; 7894/3; 7912/3; 7930/3; 7948/3; 7966/3; 7984/3; 8002/3; 8020/3; 8038/3; 8056/3; 8074/3; 8092/3; 8110/3; 8128/3; 8146/3; 8164/3; 8182/3; 8200/3; 8218/3; 8236/3; 8254/3; 8272/3; 8290/3; 8308/3; 8326/3; 8344/3; 8362/3; 8380/3; 8398/3; 8416/3; 8434/3; 8452/3; 8470/3; 8488/3; 8506/3; 8524/3; 8542/3; 8560/3; 8578/3; 8596/3; 8614/3; 8632/3; 8650/3; 8668/3; 8686/3; 8704/3; 8722/3; 8740/3; 8758/3; 8776/3; 8794/3; 8812/3; 8830/3; 8848/3; 8866/3; 8884/3; 8902/3; 8920/3; 8938/3; 8956/3; 8974/3; 8992/3; 9010/3; 9028/3; 9046/3; 9064/3; 9082/3; 9100/3; 9118/3; 9136/3; 9154/3; 9172/3; 9190/3; 9208/3; 9226/3; 9244/3; 9262/3; 9280/3; 9298/3; 9316/3; 9334/3; 9352/3; 9370/3; 9388/3; 9406/3; 9424/3; 9442/3; 9460/3; 9478/3; 9496/3; 9514/3; 9532/3; 9550/3; 9568/3; 9586/3; 9604/3; 9622/3; 9640/3; 9658/3; 9676/3; 9694/3; 9712/3; 9730/3; 9748/3; 9766/3; 9784/3; 9802/3; 9820/3; 9838/3; 9856/3; 9874/3; 9892/3; 9910/3; 9928/3; 9946/3; 9964/3; 9982/3; 10000/3; 10018/3; 10036/3; 10054/3; 10072/3; 10090/3; 10108/3; 10126/3; 10144/3; 10162/3; 10180/3; 10198/3; 10216/3; 10234/3; 10252/3; 10270/3; 10288/3; 10306/3; 10324/3; 10342/3; 10360/3; 10378/3; 10396/3; 10414/3; 10432/3; 10450/3; 10468/3; 10486/3; 10504/3; 10522/3; 10540/3; 10558/3; 10576/3; 10594/3; 10612/3; 10630/3; 10648/3; 10666/3; 10684/3; 10702/3; 10720/3; 10738/3; 10756/3; 10774/3; 10792/3; 10810/3; 10828/3; 10846/3; 10864/3; 10882/3; 10900/3; 10918/3; 10936/3; 10954/3; 10972/3; 10990/3; 11008/3; 11026/3; 11044/3; 11062/3; 11080/3; 11098/3; 11116/3; 11134/3; 11152/3; 11170/3; 11188/3; 11206/3; 11224/3; 11242/3; 11260/3; 11278/3; 11296/3; 11314/3; 11332/3; 11350/3; 11368/3; 11386/3; 11404/3; 11422/3; 11440/3; 11458/3; 11476/3; 11494/3; 11512/3; 11530/3; 11548/3; 11566/3; 11584/3; 11602/3; 11620/3; 11638/3; 11656/3; 11674/3; 11692/3; 11710/3; 11728/3; 11746/3; 11764/3; 11782/3; 11800/3; 11818/3; 11836/3; 11854/3; 11872/3; 11890/3; 11908/3; 11926/3; 11944/3; 11962/3; 11980/3; 11998/3; 12016/3; 12034/3; 12052/3; 12070/3; 12088/3; 12106/3; 12124/3; 12142/3; 12160/3; 12178/3; 12196/3; 12214/3; 12232/3; 12250/3; 12268/3; 12286/3; 12304/3; 12322/3; 12340/3; 12358/3; 12376/3; 12394/3; 12412/3; 12430/3; 12448/3; 12466/3; 12484/3; 12502/3; 12520/3; 12538/3; 12556/3; 12574/3; 12592/3; 12610/3; 12628/3; 12646/3; 12664/3; 12682/3; 12700/3; 12718/3; 12736/3; 12754/3; 12772/3; 12790/3; 12808/3; 12826/3; 12844/3; 12862/3; 12880/3; 12898/3; 12916/3; 12934/3; 12952/3; 12970/3; 12988/3; 13006/3; 13024/3; 13042/3; 13060/3; 13078/3; 13096/3; 13114/3; 13132/3; 13150/3; 13168/3; 13186/3; 13204/3; 13222/3; 13240/3; 13258/3; 13276/3; 13294/3; 13312/3; 13330/3; 13348/3; 13366/3; 13384/3; 13402/3; 13420/3; 13438/3; 13456/3; 13474/3; 13492/3; 13510/3; 13528/3; 13546/3; 13564/3; 13582/3; 13600/3; 13618/3; 13636/3; 13654/3; 13672/3; 13690/3; 13708/3; 13726/3; 13744/3; 13762/3; 13780/3; 13798/3; 13816/3; 13834/3; 13852/3; 13870/3; 13888/3; 13906/3; 13924/3; 13942/3; 13960/3; 13978/3; 13996/3; 14014/3; 14032/3; 14050/3; 14068/3; 14086/3; 14104/3; 14122/3; 14140/3; 14158/3; 14176/3; 14194/3; 14212/3; 14230/3; 14248/3; 14266/3; 14284/3; 14302/3; 14320/3; 14338/3; 14356/3; 14374/3; 14392/3; 14410/3; 14428/3; 14446/3; 14464/3; 14482/3; 14500/3; 14518/3; 14536/3; 14554/3; 14572/3; 14590/3; 14608/3; 14626/3; 14644/3; 14662/3; 14680/3; 14698/3; 14716/3; 14734/3; 14752/3; 14770/3; 14788/3; 14806/3; 14824/3; 14842/3; 14860/3; 14878/3; 14896/3; 14914/3; 14932/3; 14950/3; 14968/3; 14986/3; 15004/3; 15022/3; 15040/3; 15058/3; 15076/3; 15094/3; 15112/3; 15130/3; 15148/3; 15166/3; 15184/3; 15202/3; 15220/3; 15238/3; 15256/3; 15274/3; 15292/3; 15310/3; 15328/3; 15346/3; 15364/3; 15382/3; 15400/3; 15418/3; 15436/3; 15454/3; 15472/3; 15490/3; 15508/3; 15526/3; 15544/3; 15562/3; 15580/3; 15598/3; 15616/3; 15634/3; 15652/3; 15670/3; 15688/3; 15706/3; 15724/3; 15742/3; 15760/3; 15778/3; 15796/3; 15814/3; 15832/3; 15850/3; 15868/3; 15886/3; 15904/3; 15922/3; 15940/3; 15958/3; 15976/3; 15994/3; 16012/3; 16030/3; 16048/3; 16066/3; 16084/3; 16102/3; 16120/3; 16138/3; 16156/3; 16174/3; 16192/3; 16210/3; 16228/3; 16246/3; 16264/3; 16282/3; 16300/3; 16318/3; 16336/3; 16354/3; 16372/3; 16390/3; 16408/3; 16426/3; 16444/3; 16462/3; 16480/3; 16498/3; 16516/3; 16534/3; 16552/3; 16570/3; 16588/3; 16606/3; 16624/3; 16642/3; 16660/3; 16678/3; 16696/3; 16714/3; 16732/3; 16750/3; 16768/3; 16786/3; 16804/3; 16822/3; 16840/3; 16858/3; 16876/3; 16894/3; 16912/3; 16930/3; 16948/3; 16966/3; 16984/3; 17002/3; 17020/3; 17038/3; 17056/3; 17074/3; 17092/3; 17110/3; 17128/3; 17146/3; 17164/3; 17182/3; 17200/3; 17218/3; 17236/3; 17254/3; 17272/3; 17290/3; 17308/3; 17326/3; 17344/3; 17362/3; 17380/3; 17398/3; 17416/3; 17434/3; 17452/3; 17470/3; 17488/3; 17506/3; 17524/3; 17542/3; 17560/3; 17578/3; 17596/3; 17614/3; 17632/3; 17650/3; 17668/3; 17686/3; 17704/3; 17722/3; 17740/3; 17758/3; 17776/3; 17794/3; 17812/3; 17830/3; 17848/3; 17866/3; 17884/3; 17902/3; 17920/3; 17938/3; 17956/3; 17974/3; 17992/3; 18010/3; 18028/3; 18046/3; 18064/3; 18082/3; 18100/3; 18118/3; 18136/3; 18154/3; 18172/3; 18190/3; 18208/3; 18226/3; 18244/3; 18262/3; 18280/3; 18298/3; 18316/3; 18334/3; 18352/3; 18370/3; 18388/3; 18406/3; 18424/3; 18442/3; 18460/3; 18478/3; 18496/3; 18514/3; 18532/3; 18550/3; 18568/3; 18586/3; 18604/3; 18622/3; 18640/3; 18658/3; 18676/3; 18694/3; 18712/3; 18730/3; 18748/3; 18766/3; 18784/3; 18802/3; 18820/3; 18838/3; 18856/3; 18874/3; 18892/3; 18910/3; 18928/3; 18946/3; 18964/3; 18982/3; 19000/3; 19018/3; 19036/3; 19054/3; 19072/3; 19090/3; 19108/3; 19126/3; 19144/3; 19162/3; 19180/3; 19198/3; 19216/3; 19234/3; 19252/3; 19270/3; 19288/3; 19306/3; 19324/3; 19342/3; 19360/3; 19378/3; 19396/3; 19414/3; 19432/3; 19450/3; 19468/3; 19486/3; 19504/3; 19522/3; 19540/3; 19558/3; 19576/3; 19594/3; 19612/3; 19630/3; 19648/3; 19666/3; 19684/3; 19702/3; 19720/3; 19738/3; 19756/3; 19774/3; 19792/3; 19810/3; 19828/3; 19846/3; 19864/3; 19882/3; 19900/3; 19918/3; 19936/3; 19954/3; 19972/3; 19990/3; 20008/3; 20026/3; 20044/3; 20062/3; 20080/3; 20098/3; 20116/3; 20134/3; 20152/3; 20170/3; 20188/3; 20206/3; 20224/3; 20242/3; 20260/3; 20278/3; 20296/3; 20314/3; 20332/3; 20350/3; 20368/3; 20386/3; 20404/3; 20422/3; 20440/3; 20458/3; 20476/3; 20494/3; 20512/3; 20530/3; 20548/3; 20566/3; 20584/3; 20602/3; 20620/3; 20638/3; 20656/3; 20674/3; 20692/3; 20710/3; 20728/3; 20746/3; 20764/3; 20782/3; 20800/3; 20818/3; 20836/3; 20854/3; 20872/3; 20890/3; 20908/3; 20926/3; 20944/3; 20962/3; 20980/3; 20998/3; 21016/3; 21034/3; 21052/3; 21070/3; 21088/3; 21106/3; 21124/3; 21142/3; 21160/3; 21178/3; 21196/3; 21214/3; 21232/3; 21250/3; 21268/3; 21286/3; 21304/3; 21322/3; 21340/3; 21358/3; 21376/3; 21394/3; 21412/3; 21430/3; 21448/3; 21466/3; 21484/3; 21502/3; 21520

C;Accession: F70863  
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Felwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, K.; Skelton, S.; Squares, S.  
Nature 393, 537-544, 1998  
A;Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome  
A;Reference number: A70500; MUID:98295987; PMID:9634230  
A;Accession: F70863  
A;Status: preliminary; nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-487 <COL>  
A;Cross-references: GB:AL021246; GB:AL123456; NID:G3261507; PIDN:CAAL6024.1; PID:el123757  
A;Experimental source: strain H37RV  
C;Genetics:  
A;Gene: foic  
C;Superfamily: folylpolyglutamate synthase

Query Match 78.6%; Score 44; DB 2; Length 487;  
Best Local Similarity 90.0%; Pred. No. 12;  
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 IAGIGGEKAG 10  
|||||  
Db 209 IAGIAGEKAG 218

RESULT 6  
AE2981  
arylesterase [imported] - Agrobacterium tumefaciens (strain C58, Dupont)  
C;Species: Agrobacterium tumefaciens  
C;Date: 11-Jan-2002 #sequence\_revision 11-Jan-2002 #text\_change 18-Nov-2002  
C;Accession: AE2981  
R;Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I.; erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutayavin, T.; Levy, R.; Li, M.; McClellan; Karp, P.; Romero, P.; Zhang, S.  
Science 294, 2317-2323, 2001  
A;Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm, ster, E.W.  
A;Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.  
A;Reference number: AE2577; MUID:21608550; PMID:11743193  
A;Accession: AE2981  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-209 <KUR>  
A;Cross-references: GB:AE009689; PIDN:AAL44267.1; PID:gl7741852; GSPDB:GN00187  
A;Experimental source: strain C58 (Dupont)  
C;Genetics:  
A;Gene: ada  
A;Map position: linear chromosome

Query Match 75.0%; Score 42; DB 2; Length 209;  
Best Local Similarity 72.7%; Pred. No. 11;  
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 IAGIGGEKAG 11  
:|||||  
Db 137 VAGPGGEAG 147

RESULT 7  
arylesterase [AF044683] [imported] - Agrobacterium tumefaciens (strain C58, Cereon)  
C;Species: Agrobacterium tumefaciens  
C;Date: 22-Oct-2001 #sequence\_revision 22-Oct-2001 #text\_change 18-Nov-2002  
C;Accession: C98302  
R;Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Quorollo, B.; Goldman, A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.; Science 294, 2323-2328, 2001  
A;Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tumefaciens C58  
A;Reference number: A97359; MUID:21608551; PMID:11743194  
A;Accession: C98302  
A;Status: preliminary

A;Molecule type: DNA  
A;Residues: 1-209 <KUR>  
A;Cross-references: GB:AE007870; PIDN:AAK89941.1; PID:gl5159896; GSPDB:GN00170  
C;Genetics:  
A;Gene: AGR L 2749  
A;Map position: linear chromosome

Query Match 75.0%; Score 42; DB 2; Length 209;  
Best Local Similarity 72.7%; Pred. No. 11;  
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 IAGIGGEKAG 11  
:|||||  
Db 137 VAGPGGEAG 147

RESULT 8  
F69732  
PBSX prophage ORF xkdo - Bacillus subtilis  
C;Species: Bacillus subtilis  
C;Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 15-Oct-1999  
C;Accession: F69732  
R;Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berter C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Cho A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E. Nature 390, 249-256, 1997  
A;Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galler lech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hulio, M.F. Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois, A;Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel Y. M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon, A;Authors: Schleib, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seror akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terptrata, P.; Tognoni, A.; Tosato, V.; Uchiyama, T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K A;Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.  
A;Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.  
A;Reference number: A69580; MUID:98044033; PMID:9384377  
A;Accession: F69732  
A;Status: preliminary; nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-1332 <KUN>  
A;Cross-references: GB:Z99110; GB:AL009126; NID:G2633472; PIDN:CAB13125.1; PID:el183288; A;Experimental source: strain 168  
C;Genetics:  
A;Gene: xkdo

Query Match 73.2%; Score 41; DB 2; Length 1332;  
Best Local Similarity 70.0%; Pred. No. 92;  
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 IAGIGGEKAG 10  
:|||||  
Db 469 VGGIGGEAG 478

RESULT 9  
H69074  
formylmethanofuran dehydrogenase (EC 1.2.99.5) (tungsten) chain C - Methanobacterium ther C;Species: Methanobacterium thermoautotrophicum  
C;Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 24-Sep-1999  
C;Accession: H69074  
R;Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.; Dubois, J.; Aldredge, T.; Qiu, D.; Spadafora, R.; Vicaire, R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiwani, N.; Ki, S.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N. J. Bacteriol. 179, 7135-7155, 1997  
A;Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta H: funct A;Reference number: A69000; MUID:98037514; PMID:9371463  
A;Accession: H69074  
A;Status: preliminary; nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-270 <MTH>  
A;Cross-references: GB:AB000916; GB:AE000666; NID:G2622674; PIDN:AAH86032.1; PID:G262268

```
A:Experimental source: strain Delta H
C:Genetics:
A:Gene: MTH1558
C:Superfamily: formylmethanofuran dehydrogenase chain C
C:Keywords: iron-sulfur protein; metalloprotein; oxidoreductase; tungsten

Query Match 71.4%; Score 40; DB 2; Length 270;
Best Local Similarity 63.6%; Pred. No. 31;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 IAGIGGERAGG 11
Db 197 VARVGGEWAGG 207

RESULT 10
S57457
formylmethanofuran dehydrogenase (EC 1.2.99.5) (tungsten) chain C - Methanobacterium the
C:Species: Methanobacterium thermoautotrophicum
C:Date: 10-Oct-1995 #sequence_revision 17-Nov-1995 #text_change 20-Jun-2000
C:Accession: S63546; S57457
R:Hochheimer, A.; Schmitz, R.A.; Thauer, R.K.; Hedderich, R.
Eur. J. Biochem. 234, 910-920, 1995
A:Title: The tungsten formylmethanofuran dehydrogenase from Methanobacterium thermoautot
A:Reference number: S63519; MUID:96163477; PMID:8575452
A:Accession: S63546
A:Molecule type: DNA
A:Residues: 1-270 <HOC>
A:Cross-references: EMBL:X87970; NID:g1890205; PIDN:CAA61214.1; PID:g1890211
A:Experimental source: strain Marburg, DSM 2133
C:Genetics:
A:Gene: fwdC
A:Complex: heterotetramer of A (see PIR:S57456), B (see PIR:S57458), C, and D (see PIR:S
C:Superfamily: formylmethanofuran dehydrogenase chain C
C:Keywords: heterotetramer; iron-sulfur protein; metalloprotein; oxidoreductase; tungste

Query Match 71.4%; Score 40; DB 2; Length 270;
Best Local Similarity 63.6%; Pred. No. 31;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 IAGIGGERAGG 11
Db 197 VARVGGEWAGG 207

RESULT 11
E95286
hypothetical protein SMO367 [imported] - Sinorhizobium meliloti (strain 1021) magaplas
C:Species: Sinorhizobium meliloti
C:Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 30-Sep-2001
C:Accession: E95286
R:Barrett, M.J.; Fisher, R.F.; Jones, T.; Komp, C.; Abola, A.P.; Barloy-Hubler, F.; Bows
.; Kalman, S.; Keating, D.H.; Palm, C.; Peck, M.C.; Surzycki, R.; Wells, D.H.; Yeh, K.C.
Proc. Natl. Acad. Sci. U.S.A. 98, 9883-9888, 2001
A:Title: Nucleotide sequence and predicted functions of the entire Sinorhizobium melilot
A:Reference number: A95262; MUID:21396509; PMID:11481432
A:Accession: E95286
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-298 <KUR>
A:Cross-references: GB:AE006469; PIDN:AK64855.1; PID:g14523269; GSPDB:GN00165
A:Experimental source: strain 1021, megaplasmid pSymA
R:Galibert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubler,
pella, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.;
L.; Hyman, R.W.; Jones, T.
Science 293, 668-672, 2001
A:Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelaure,
neault, P.; Vandenberg, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K.
A:Title: The composite genome of the legume symbiont Sinorhizobium meliloti.
A:Reference number: A96039; MUID:21368234; PMID:11474104
C:Contents: annotation
C:Genetics:
A:Gene: SMO367

A:Genome: plasmid
Query Match 71.4%; Score 40; DB 2; Length 298;
Best Local Similarity 88.9%; Pred. No. 34;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AGIGGEKAG 10
Db 103 AGIGGAKAG 111

RESULT 12
TS1035
hypothetical protein B15120.50 [imported] - Neurospora crassa
C:Species: Neurospora crassa
C:Date: 21-Jul-2000 #sequence_revision 21-Jul-2000 #text_change 21-Jul-2000
C:Accession: TS1035
R:Schulte, U.; Aign, V.; Hoheisel, J.; Brandt, P.; Partmann, B.; Holland, R.; Nyakatura,
submitted to the Protein Sequence Database, July 2000
A:Reference number: Z5286
A:Accession: TS1035
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-548 <SCH>
A:Cross-references: EMBL:AL389900; GSPDB:GN00116; NCSP:B15120.50
A:Experimental source: BAC clone B15120; Strain OR74A
C:Genetics:
A:Gene: NCSP:B15120.50
A:Map position: 6

Query Match 71.4%; Score 40; DB 2; Length 548;
Best Local Similarity 63.6%; Pred. No. 59;
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 IAGIGGERAGG 11
Db 294 VAGAGENEKG 304

RESULT 13
T14302
glycine-rich cell wall protein - carrot (fragment)
C:Species: Daucus carota (carrot)
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
C:Accession: T14302
R:Lin, X.; Hwang, G.J.; Zimmerman, J.L.
submitted to the EMBL Data Library, January 1996
A:Description: Isolation and characterization of a diverse set of genes from carrot somat
A:Reference number: Z17968
A:Accession: T14302
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-87 <LIN>
A:Cross-references: EMBL:U47088; NID:g1276957; PID:g1276958
A:Experimental source: strain Danver Half-long
C:Keywords: cell wall

Query Match 69.6%; Score 39; DB 2; Length 87;
Best Local Similarity 63.6%; Pred. No. 16;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 IAGIGGERAGG 11
Db 68 LAGVGGVAGG 78

RESULT 14
T47394
hypothetical protein T18D12.90 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cross)
C:Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 20-Apr-2000
C:Accession: T47394
R:Massenet, O.; Clabault, G.; Quigley, F.; Mache, R.; Mewes, H.W.; Lemcke, K.; Mayer, K.I
```

submitted to the Protein Sequence Database, March 2000

A;Reference number: Z24464  
A;Accession: I47394  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-163 <MAS>  
A;Cross-references: EMBL:AL138644  
A;Experimental source: cultivar Columbia; BAC clone T18D12  
C;Genetics:  
A;Map position: 3  
A;Introns: 150/1  
A;Note: T18D12.90

Query Match 69.6%; Score 39; DB 2; Length 163;  
Best Local Similarity 77.8%; Pred. No. 28;  
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 3 GIGGEKAGG 11  
|||:|  
Db 99 GIGGDKFEG 107

RESULT 15  
I49754  
homeobox protein - mouse  
N;Alternate names: homeotic protein MAB66  
C;Species: Mus musculus (house mouse)  
C;Date: 29-May-1998 #sequence\_revision 29-May-1998 #text\_change 17-Nov-2000  
C;Accession: I49754; F37290; B41606; F38809  
R;Benson, G.V.; Nguyen, T.H.; Maas, R.L.  
Mol. Cell. Biol. 15, 1591-1601, 1995  
A;Title: The expression pattern of the murine Hoxa-10 gene and the sequence recognition  
A;Reference number: I49754; MUID:95166244; PMID:7862151  
A;Accession: I49754  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: mRNA  
A;Residues: 1-399 <RES>  
A;Cross-references: GB:I08757; NID:9825647; PIDN:AAA67125.1; PID:9567213  
R;Singh, G.; Kaur, S.; Stock, J.L.; Jenkins, N.A.; Gilbert, D.J.; Copeland, N.G.; Potter  
Proc. Natl. Acad. Sci. U.S.A. 88, 10706-10710, 1991  
A;Title: Identification of 10 murine homeobox genes.  
A;Reference number: A37290; MUID:92073356; PMID:1683707  
A;Accession: F37290  
A;Status: preliminary; not compared with conceptual translation  
A;Molecule type: DNA  
A;Residues: 325-384 <SIN>  
R;Murtha, M.T.; Leckman, J.F.; Ruddle, F.H.  
Proc. Natl. Acad. Sci. U.S.A. 88, 10711-10715, 1991  
A;Title: Detection of homeobox genes in development and evolution.  
A;Reference number: A41606; MUID:92073357; PMID:1720547  
A;Accession: B41606  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 346-370 <MUR>  
A;Cross-references: GB:M81659; NID:gl193973; PIDN:AAA63312.1; PID:gl193974  
C;Genetics:  
A;Gene: Hoxa10  
C;Superfamily: unassigned homeobox proteins; homeobox homology  
C;Keywords: DNA binding; homeobox; nucleus; transcription regulation  
F;326-382/Domain: homeobox homology <HOX>

Query Match 69.6%; Score 39; DB 2; Length 399;  
Best Local Similarity 70.0%; Pred. No. 64;  
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 AGIGGEKAGG 11  
|||:|  
Db 25 AGVGGSAGG 34

Search completed: September 18, 2004, 04:30:53  
Job time : 15.3725 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 18, 2004, 03:48:54 ; Search time 8.19608 Seconds  
(without alignments)  
69.884 Million cell updates/sec

Title: US-10-615-959-42  
Perfect score: 56  
Sequence: 1 IAGIGGEKAGG 11

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_42.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	56	100.0	1466	1 CA13 HUMAN	P02461 homo sapien
2	52	92.9	636	1 CA13 RAT	P13941 rattus norv
3	52	92.9	1464	1 CA13 MOUSE	P08121 mus musculu
4	44	78.6	161	1 HS21 MAIZE	P24631 zea mays (m
5	41	73.2	1332	1 XKDO BACSU	P54334 bacillus su
6	40	71.4	270	1 FWDC METTH	O27600 methanobact
7	40	71.4	270	1 FWDC METTM	O59579 methanobact
8	39	69.6	378	1 CSP FLACL	P08675 plasmodium
9	39	69.6	399	1 HXAA MOUSE	P31310 mus musculu
10	39	69.6	1679	1 FUR2 DROME	P30432 drosophila
11	38	67.9	423	1 TBX2 CAEEL	Q19691 caenorhabdi
12	38	67.9	495	1 AMPA PSEAE	O68822 pseudomonas
13	37	66.1	269	1 POLG PWMI	P32574 passionfrui
14	37	66.1	354	1 RFL LEPIN	Q8f956 leptospira
15	37	66.1	371	1 TRMT YERPE	Q8zf55 yersinia pe
16	37	66.1	462	1 TRPE YETHE	P05378 thermus the
17	37	66.1	499	1 DLDH YEAST	P09624 saccharomyc
18	37	66.1	562	1 NPMR BACME	Q00891 bacillus me
19	37	66.1	1113	1 COR1 MOUSE	Q92319 mus musculu
20	36	64.3	214	1 YDL8 SCHPO	P87124 schizosacch
21	36	64.3	264	1 RS3 FALSO	Q8xv18 raltstonia s
22	36	64.3	270	1 FWDC METWO	O74031 methanobact
23	36	64.3	301	1 FMDC METBA	Q48943 methanosarc
24	36	64.3	330	1 ILVC CHLTE	Q8ker7 chlorobium
25	36	64.3	331	1 ILVC THETN	Q8rdk4 thermoanaer
26	36	64.3	365	1 FL3H DIACA	Q05964 dianthus ca
27	36	64.3	368	1 YJHT ECOLI	P39371 escherichia
28	36	64.3	386	1 RPA2 METVA	P14247 methanococc
29	36	64.3	460	1 MB48 MYCTU	Q933k8 mycobacteri
30	36	64.3	461	1 DYL SCHPO	P36591 schizosacch
31	36	64.3	649	1 HS70 PARBR	P87047 paracoccidi
32	36	64.3	1067	1 SGG DROME	P18431 drosophila
33	36	64.3	1137	1 RIR1 HSV11	P08543 herpes simp

34 36 64.3 1156 1 GLH4 CAEEL 075743 caenorhabdi  
35 35 62.5 147 1 RS12\_PYRAB Q9v110 pyrococcus  
36 35 62.5 147 1 RS12\_PYRHO OS9229 pyrococcus  
37 35 62.5 147 1 RS12\_THECE P29161 thermococcu  
38 35 62.5 155 1 DHB CRAPL P22239 craterostig  
39 35 62.5 312 1 YD63 THEAC Q9h1h6 thermoplasm  
40 35 62.5 373 1 TRMU\_STRMU O8drs4 streptococc  
41 35 62.5 460 1 FD3C\_RICCO P48619 ricinus com  
42 35 62.5 559 1 MERA\_ALCSP P94188 alcaligenes  
43 35 62.5 591 1 POXB\_STRPN Q54970 streptococc  
44 35 62.5 592 1 RB56 HUMAN Q92804 homo sapien  
45 35 62.5 697 1 PURL\_PYRAE Q8zzj7 pyrobaculum

## ALIGNMENTS

RESULT 1  
CA13 HUMAN  
ID CA13 HUMAN STANDARD; PRT; 1466 AA.  
AC P02461; Q15112;  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 01-JAN-1990 (Rel. 13, Last sequence update)  
DE 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Collagen alpha 1(III) chain precursor.  
GN COL3A1.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Skin fibroblast;  
RX MEDLINE=89350838; PubMed=2764886;  
RA Ala-Kokko L., Kontusaari S., Baldwin C.T., Kuivaniemi H.,  
RA Prockop D.J.;  
RT "Structure of cDNA clones coding for the entire prepro alpha 1 (III)  
RT chain of human type III procollagen. Differences in protein structure  
RT from type I procollagen and conservation of codon preferences.";  
RL Biochem. J. 260:509-516(1989).  
RN [2]  
RP SEQUENCE OF 149-1225 FROM N.A.  
RX MEDLINE=89386015; PubMed=2780304;  
RA Janeczko R.A., Ramirez F.;  
RT "Nucleotide and amino acid sequences of the entire human alpha 1  
RT (III) collagen.";  
RL Nucleic Acids Res. 17:6742-6742(1989).  
RN [3]  
RP SEQUENCE OF 168-398.  
RX MEDLINE=77134724; PubMed=557335;  
RA Seyer J.M., Kang A.H.;  
RT "Covalent structure of collagen: amino acid sequence of cyanogen  
RT bromide peptides from the amino-terminal segment of type III collagen  
RT of human liver.";  
RL Biochemistry 16:1158-1164(1977).  
RN [4]  
RP REVISIONS.  
RA Seyer J.M.;  
RT Submitted (DEC-1977) to the PIR data bank.  
RN [5]  
RP SEQUENCE OF 399-727.  
RX MEDLINE=79000343; PubMed=687591;  
RA Seyer J.M., Kang A.H.;  
RT "Covalent structure of collagen: amino acid sequence of five  
RT consecutive CNBr peptides from type III collagen of human liver.";  
RL Biochemistry 17:3404-3411(1978).  
RN [6]  
RP SEQUENCE OF 728-964.  
RX MEDLINE=80198282; PubMed=6246925;  
RA Seyer J.M., Mainardi C., Kang A.H.;  
RT "Covalent structure of collagen: amino acid sequence of alpha 1  
RT (III)-C95 from type III collagen of human liver.";  
RL Biochemistry 19:1583-1589(1980).



RA SEQUENCE OF 950-1466 FROM N.A.  
RP MEDLINE=88189827; PubMed=3357782;  
RX Mankoo B.S., Daigleish R.;  
RA "Human pro alpha 1(III) collagen: cDNA sequence for the 3' end.";  
RL Nucleic Acids Res. 16:2337-2337(1988).  
RN [8]  
RP REVISION TO 1184.  
RX MEDLINE=89098346; PubMed=3211760;  
RA Molynaux K., Daigleish R.;  
RT "Human type III collagen 'variant' is a cDNA cloning artefact.";  
RL Nucleic Acids Res. 16:11833-11833(1988).  
RN [9]  
RP SEQUENCE OF 1065-1466 FROM N.A.  
RX MEDLINE=85087944; PubMed=6096827;  
RA Loidl H.R., Brinker J.W., May M., Pihlajaniemi T., Morrow S.,  
RX Rosenbloom J., Myers J.C.;  
RT "Molecular cloning and carboxyl-propeptide analysis of human type III  
procollagen.";  
RL Nucleic Acids Res. 12:9383-9394(1984).  
RN [10]  
RP SEQUENCE OF 965-1200.  
RX MEDLINE=81208139; PubMed=7016180;  
RA Seyer J.M., Kang A.H.;  
RT "Covalent structure of collagen: amino acid sequence of alpha  
1(III)-CB9 from type III collagen of human liver.";  
RL Biochemistry 20:2621-2627(1981).  
RN [11]  
RP SEQUENCE OF 1176-1466 FROM N.A.  
RX MEDLINE=85157600; PubMed=2579949;  
RA Chu M.-L., Weil D., de Wet W.J., Bernard M.P., Sippola M., Ramirez F.;  
RT "Isolation of cDNA and genomic clones encoding human pro-alpha 1  
(III) collagen. Partial characterization of the 3' end region of the  
gene.";  
RL J. Biol. Chem. 260:4357-4363(1985).  
RN [12]  
RP SEQUENCE OF 1161-1200 FROM N.A.  
RX MEDLINE=86187804; PubMed=3754462;  
RA Miskulin M., Daigleish R., Klueve-Beckerman B., Rennard S.I.,  
RA Tolstoshev P., Brantly M., Crystal R.G.;  
RT "Human type III collagen gene expression is coordinately modulated  
with the type I collagen genes during fibroblast growth.";  
RL Biochemistry 25:1408-1413(1986).  
RN [13]  
RP SEQUENCE OF 1-170 FROM N.A.  
RC TISSUE-Placenta; PubMed=3405773;  
RX MEDLINE=88303360; PubMed=3405773;  
RA Toman D., Ricca G., de Crombrughe B.;  
RT "Nucleotide sequence of a cDNA coding for the amino-terminal region  
of human prepro alpha 1(III) collagen.";  
RL Nucleic Acids Res. 16:7201-7201(1988).  
RN [14]  
RP SEQUENCE OF 1-176 FROM N.A.  
RX MEDLINE=89378752; PubMed=2777083;  
RA Benson-Chanda V., Su M.W., Weil D., Chu M.-L., Ramirez F.;  
RT "Cloning and analysis of the 5' portion of the human type-III  
procollagen gene (COL3A1).";  
RL Gene 78:255-265(1989).  
RN [15]  
RP REVIEW ON VARIANTS.  
RX MEDLINE=97255959; PubMed=9101290;  
RA Kuivaniemi H., Tromp G., Prockop D.J.;  
RT "Mutations in fibrillar collagen (types I, II, III, and XI), fibril-  
associated collagen (type IX), and network-forming collagen (type X)  
cause a spectrum of diseases of bone, cartilage, and blood vessels.";  
RL Hum. Mutat. 9:300-315(1997).  
RN [16]  
RP VARIANT AORTIC ANEURYSM ARG-303, AND VARIANT THR-668.  
RX MEDLINE=93293988; PubMed=8514866;  
RA Tromp G., Wu Y., Prockop D.J., Madhaderi S.L., Kleinert C.,  
RA Earley J.J., Zhuang J., Noerregaard O., Darling R.C., Abbott W.M.,  
RA Cole C.W., Jaakkola P., Rynnaenen M., Pearce W.H., Yao J.S.T.,  
RA Majamaa K., Smullens S.N., Gatalica Z., Ferrell R.E., Jimenez S.A.,  
RA Jackson C.E., Michels V.V., Kaye M., Kuivaniemi H.;  
RT "Sequencing of cDNA from 50 unrelated patients reveals that mutations  
in the triple-helical domain of type III procollagen are an  
independent cause of aortic aneurysms.";  
RL J. Clin. Invest. 91:2539-2545(1993).  
RN [17]  
RP VARIANT THR-698.  
RX MEDLINE=91045136; PubMed=2235526;  
RA Zafarullah K., Kleinert C., Tromp G., Kuivaniemi H., Kontusaari S.,  
RA Wu Y., Ganguly A., Prockop D.J.;  
RT "G to A polymorphism in exon 31 of the COL3A1 gene.";  
RL Nucleic Acids Res. 18:6180-6180(1990).  
RN [18]  
RP VARIANT AORTIC ANEURYSM ARG-786.  
RX MEDLINE=91056145; PubMed=2243125;  
RA Kontusaari S., Tromp G., Kuivaniemi H., Romanic A.M., Prockop D.J.;  
RT "A mutation in the gene for type III procollagen (COL3A1) in a family  
with aortic aneurysms.";  
RL J. Clin. Invest. 86:1465-1473(1990).  
RN [19]  
RP VARIANT EDS-IV ARG-828.  
RX MEDLINE=94016385; PubMed=8411057;  
RA Richards A.J., Narcisi P., Lloyd J.C., Ferguson C., Pope F.M.;  
RT "The substitution of glycine 661 by arginine in type III collagen  
produces mutant molecules with different thermal stabilities and  
causes Ehlers-Danlos syndrome type IV.";  
RL J. Med. Genet. 30:690-693(1993).  
RN [20]  
RP VARIANT EDS-IV SER-957.  
RX MEDLINE=89109135; PubMed=2492273;  
RA Tromp G., Kuivaniemi H., Shikata H., Prockop D.J.;  
RT "A single base mutation that substitutes serine for glycine 790 of  
the alpha 1 (III) chain of type III procollagen exposes an arginine  
and causes Ehlers-Danlos syndrome IV.";  
RL J. Biol. Chem. 264:1349-1352(1989).  
RN [21]  
RP VARIANT EDS-IV VAL-960.  
RX MEDLINE=95288429; PubMed=7749417;  
RA Tromp G., de Paeppe A., Nuytink L., Madhaderi S.L., Kuivaniemi H.;  
RT "Substitution of valine for glycine 793 in type III procollagen in  
Ehlers-Danlos syndrome type IV.";  
RL Hum. Mutat. 5:179-181(1995).  
RN [22]  
RP VARIANT EDS-IV GLU-1014.  
RX MEDLINE=92316511; PubMed=1352273;  
RA Richards A.J., Ward P.N., Narcisi P., Nicholls A.C., Lloyd J.C.,  
RA Pope F.M.;  
RT "A single base mutation in the gene for type III collagen (COL3A1)  
converts glycine 847 to glutamic acid in a family with Ehlers-Danlos  
syndrome type IV. An unaffected family member is mosaic for the  
mutation.";  
RL Hum. Genet. 89:414-418(1992).  
RN [23]  
RP VARIANT EDS-IV ASP-1050.  
RX MEDLINE=90037070; PubMed=2808425;  
RA Tromp G., Kuivaniemi H., Stolle C.A., Pope F.M., Prockop D.J.;  
RT "Single base mutation in the type III procollagen gene that converts  
the codon for glycine 883 to aspartate in a mild variant of  
Ehlers-Danlos syndrome IV.";  
RL J. Biol. Chem. 264:19313-19317(1989).  
RN [24]  
RP VARIANT EDS-IV VAL-1077.  
RX MEDLINE=91374480; PubMed=1895316;  
RA Richards A.J., Lloyd J.C., Ward P.N., de Paeppe A., Narcisi P.,  
RA Pope F.M.;  
RT "Characterisation of a glycine to valine substitution at amino acid  
position 910 of the triple helical region of type III collagen in a  
patient with Ehlers-Danlos syndrome type IV.";  
RL J. Med. Genet. 28:458-463(1991).  
RN [25]  
RP VARIANT EDS-IV GLU-1173.  
RX MEDLINE=93022543; PubMed=1357232;  
RA Johnson P.H., Richards A.J., Pope F.M., Hopkinson D.A.;



DR	SMART; SMO0038; COLFI; 1.	
DR	PROSITE; PS01208; WVCF_1; PARTIAL.	
KW	Extracellular matrix; Connective tissue; Repeat; Hydroxylation;	
DR	Collagen; Glycoprotein.	
FT	NON_TER 1	
FT	CHAIN <1 375	COLLAGEN ALPHA 1(III) CHAIN.
FT	PROPEP 376	CARBOXYL-TERMINAL PROPEPTIDE.
FT	DOMAIN <1 636	TRIPLE-HELICAL REGION.
FT	DOMAIN 369 636	NONHELICAL REGION (C-TERMINAL).
FT	DISULFID 368 368	INTERCHAIN (BY SIMILARITY).
FT	DISULFID 369 369	INTERCHAIN (BY SIMILARITY).
FT	CONFLICT 340 340	N -> D (IN REF. 2).
FT	CONFLICT 429 429	A -> G (IN REF. 2).
FT	SEQUENCE 636 AA; 62332 MW; 61A48159F01D01EE CRC64;	
QY	Query Match 92.9%; Score 52; DB 1; Length 636;	
DB	Best Local Similarity 81.8%; Pred. No. 0.43;	
	Matches 9; Conservative 2; Mismatches 0; Indels 0; Gaps	
QY	1 IAGIGGEXKAGG 11	
DB	375 IAGVGGEKSGG 385	
RESULT 3		
ID	CA13_MOUSE	STANDARD; PRT; 1464 AA.
AC	P08121; Q61429; Q9CRN7;	
DC	01-AUG-1988 (Rel. 08, Created)	
DT	15-JUL-1999 (Rel. 38, Last sequence update)	
DE	15-MAR-2004 (Rel. 43, Last annotation update)	
DE	Collagen alpha 1(III) chain precursor.	
GN	COL3A1.	
OS	Mus musculus (Mouse).	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.	
OX	NCBI_TaxID=10090;	
RP	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=C57BL/6 X DBA; TISSUE=Embryo;	
RC	MEDLINE=95011609; PubMed=7926795;	
RA	Toman D., de Crombrughe B.;	
RT	"The mouse type-III procollagen-encoding gene: genomic cloning and	
RT	complete DNA sequence."	
RL	Gene 147:161-168(1994).	
RL	[2]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=C57BL/6; TISSUE=Brain;	
RC	MEDLINE=22388257; PubMed=12477932;	
RA	Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,	
RA	Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,	
RA	Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,	
RA	Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,	
RA	Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,	
RA	Scapleton M., Soares M.B., Donald M.F., Casavant T.L., Scheetz T.E.,	
RA	Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,	
RA	Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,	
RA	Bosak S.A., McSwain P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,	
RA	Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,	
RA	Villalon D.K., Munzy D.W., Sodergren E.J., Lu X., Gibbs R.A.,	
RA	Fahy J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,	
RA	Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,	
RA	Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,	
RA	Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,	
RA	Butterfield Y.S.N., Krzywiniski M.I., Skalska U., Smailus D.E.,	
RA	Schmurch A., Schein J.E., Jones S.J.M., Marra M.A.;	
RT	"Generation and initial analysis of more than 15,000 full-length	
RT	human and mouse cDNA sequences."	
RL	Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).	
RL	[3]	
RP	SEQUENCE OF 1-488 FROM N.A.	
RX	MEDLINE=88167858; PubMed=3443309;	
RX	Wood L., Thériault N., Vogeli G.;	

"Complete nucleotide sequence of the N-terminal domains of the murine alpha-1 type-III collagen chain.";  
Gene 61:225-230(1987).  
[4]

SEQUENCE OF 1-28 FROM N.A.  
MEDLINE=85131189; PubMed=3972847;  
Liau G., Mudryj M., de Crombrughe B.;  
"Identification of the promoter and first exon of the mouse alpha 1 (III) collagen gene.";  
J. Biol. Chem. 260:3773-3777(1985).  
[5]

SEQUENCE OF 810-1464 FROM N.A.  
STRAIN=C57BL/6J; TISSUE=Embryonic head;  
MEDLINE=21085660; PubMed=11217851;  
Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y., Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S., Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaoka I., Saito T., Okazaki Y., Gojohori T., Bono H., Kasukawa T., Saito R., Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T., Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H., Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J., Schirl L.M., Straubli F., Suzuki R., Tomita M., Wagner L., Washio T., Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G., Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F., Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M., Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H., Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P., Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N., Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F., Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L., Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S., Hayashizaki Y.;  
"Functional annotation of a full-length mouse cDNA collection.";  
Nature 409:685-690(2001).  
[6]

SEQUENCE OF 1442-1464 FROM N.A.  
STRAIN=C57BL/6J;  
MEDLINE=91274355; PubMed=2054384;  
Metsaeranta M., Toman D., de Crombrughe B., Vuorio E.;  
"Specific hybridization probes for mouse type I, II, III and IX collagen mRNAs.";  
Biochim. Biophys. Acta 1089:241-243(1991).

-I- FUNCTION: Collagen type III occurs in most soft connective tissues along with type I collagen.  
-I- SUBUNIT: Trimers of identical alpha 1(III) chains. The chains are linked to each other by interchain disulfide bonds. Trimers are also cross-linked via hydroxylysines.  
-I- PTM: Proline residues at the third position of the tripeptide repeating unit (G-X-Y) are hydroxylated in some or all of the chains.  
-I- PTM: O-linked glycan consists of a Glc-Gal disaccharide bound to the oxygen atom of a post-translationally added hydroxyl group (By similarity).  
-I- SIMILARITY: Contains 1 VWF domain.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.ebi.ac.uk/announcements> or send an email to [license@ebi.ac.uk](mailto:license@ebi.ac.uk)).

EMBL; X52046; CAA36279.1; -;  
EMBL; BC043089; AAH43089.1; -;  
EMBL; BC058724; AAH58724.1; -;  
EMBL; M18933; AAA37338.1; -;  
EMBL; K03037; -; NOT ANNOTATED\_CDS.  
EMBL; AK019448; BAB31724.1; -;  
EMBL; X57983; CAA41048.1; -;  
PIR; A27353; A27353.  
PIR; S59856; S59856.  
MGD; MGI:88453; Col3a1.

DR InterPro; IPR008161; C1g helix.  
DR InterPro; IPR008160; Collagen.  
DR InterPro; IPR00885; Fib collagen C.  
DR InterPro; IPR02181; Fibrinogen C-  
DR Pfam; PF01410; COLFI; 1.  
DR Pfam; PF01391; Collagen; 18.  
DR ProDom; PD000007; C1g helix; 1.  
DR ProDom; PD002078; Fib collagen\_C; 1.  
DR SMART; SM00038; COLFI; 1.  
DR SMART; SM00214; VWC; 1.  
DR PROSITE; PS01208; VWF; 1; 1.  
DR PROSITE; PS50184; VWF-2; 1.  
KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;  
KW Glycoprotein; Collagen; Signal.  
FT SIGNAL 1 23 BY SIMILARITY.  
FT PROPEP 24 154 AMINO-TERMINAL PROPEPTIDE.  
FT CHAIN 155 1203 COLLAGEN ALPHA 1(III) CHAIN.  
FT PROPEP 1204 1464 CARBOXYL-TERMINAL PROPEPTIDE.  
FT DOMAIN 31 90 VWF.  
FT DOMAIN 155 169 NONHELICAL REGION (N-TERMINAL).  
FT DOMAIN 170 1195 TRIPLE-HELICAL REGION.  
FT DOMAIN 1196 1464 NONHELICAL REGION (C-TERMINAL).  
FT CARBOHYD 262 262 O-LINKED (GAL...) (BY SIMILARITY).  
FT MOD\_RES 262 262 HYDROXYLATION (BY SIMILARITY).  
FT MOD\_RES 283 283 HYDROXYLATION (BY SIMILARITY).  
FT MOD\_RES 859 859 HYDROXYLATION (BY SIMILARITY).  
FT MOD\_RES 976 976 HYDROXYLATION (BY SIMILARITY).  
FT MOD\_RES 1093 1093 HYDROXYLATION (BY SIMILARITY).  
FT MOD\_RES 1105 1105 HYDROXYLATION (BY SIMILARITY).  
FT DISULFID 1195 1195 INTERCHAIN (BY SIMILARITY).  
FT DISULFID 1196 1196 INTERCHAIN (BY SIMILARITY).  
SQ SEQUENCE 1464 AA; 138944 MW; 2104EC27A886090B CRC64;

Query Match 92.9%; Score 52; DB 1; Length 1464;

Best Local Similarity 81.8%; Pred. No. 0.92;

Matches 9; Conservative

Mismatches 2; Indels 0; Gaps 0;

Qy 1 IAGIGGKAGG 11

Db 1203 IAGVGGEKSGG 1213

|||||:|||||

|||||:|||||

|||||:|||||

|||||:|||||

|||||:|||||

|||||:|||||

|||||:|||||

|||||:|||||

|||||:|||||

|||||:|||||

|||||:|||||

|||||:|||||

|||||:|||||

|||||:|||||

|||||:|||||

|||||:|||||

|||||:|||||

|||||:|||||

|||||:|||||

|||||:|||||

|||||:|||||

|||||:|||||

|||||:|||||

|||||:|||||

|||||:|||||

|||||:|||||

|||||:|||||

|||||:|||||

|||||:|||||

|||||:|||||

|||||:|||||

|||||:|||||

|||||:|||||

|||||:|||||

|||||:|||||

|||||:|||||

|||||:|||||

|||||:|||||

|||||:|||||

|||||:|||||

|||||:|||||

|||||:|||||

|||||:|||||

|||||:|||||

|||||:|||||

|||||:|||||

|||||:|||||

|||||:|||||

|||||:|||||

|||||:|||||

|||||:|||||

|||||:|||||

|||||:|||||

```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X54076; CAA38013.1; -.
DR PIR; S14998; S14998.
DR MaizeDB; 51309; -.
DR InterPro; IPR002068; Hsp20.
DR InterPro; IPR008978; HSP20_chap.
DR Pfam; PF00011; HSP20; 1.
DR PROSITE; PS01031; HSP20; 1.
KW Heat shock; Multigene family.
SQ SEQUENCE 161 AA; 17511 MW; 3E60A8DD4396577C CRC64;

Query Match 78.6%; Score 44; DB 1; Length 161;
Best Local Similarity 80.0%; Pred. No. 2.4;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 AGIGGEKAGG 11
|||:|||||
DB 27 AGAGGDKAGG 36

RESULT 5
XKDO_BACSU STANDARD; PRT; 1332 AA.
ID PK54334
AC Q27650
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Phage-like element PBSX protein xkdo.
GN XKDO OR BSU12680.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN STRAIN=168;
RP SEQUENCE FROM N.A.
RA Krogh S., O'Reilly M., Nolan N., Devine K.M.;
RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Barbero M.G., Bessieres P., Bolotin A., Borchert S.,
RA Boriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Friez C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Ghim S.F., Glaser P., Goffeau A., Goughly E.J., Grandi G.,
RA Guisepi G., Guy B.J., Haga K., Haeche J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Maeda S., Maue C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moesti D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogilwa A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portetelle D., Porwollig S., Prescott A.M.,
RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA Sekiguchi J., Sekowska A., Seror S.J., Serro P., Shin B.S., Soldo B.,
RA Sorokin A., Taconi E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzenecker T.,

```

```

RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;
RT "The complete genome sequence of the Gram-positive bacterium Bacillus
RT subtilis.";
RL Nature 390:249-256(1997).
CC -!- SIMILARITY: STRONG, TO B.SUBTILIS YOBO.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z70177; CAA94037.1; -.
DR EMBL; Z99110; CAB13125.1; -.
DR PIR; F69732; F69732.
DR SubtilList; BGI1549; xkdo.
DR InterPro; IPR008258; SLT_dom.
DR Pfam; PF01464; SLT; 1.
KW Complete proteome.
SQ SEQUENCE 1332 AA; 145149 MW; A96C9C9F9E31DF01 CRC64;

Query Match 73.2%; Score 41; DB 1; Length 1332;
Best Local Similarity 70.0%; Pred. No. 51;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 IAGIGGEKAG 10
:|||||:|
DB 469 VGGIGGEKAG 478

RESULT 6
FWDC_METTH STANDARD; PRT; 270 AA.
ID Q27650
AC Q27650
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Tungsten-containing formylmethanofuran dehydrogenase II subunit C
DE (EC 1.2.99.5).
GN FWDC OR MTH1558.
OS Methanobacterium thermoautotrophicum.
OC Archaea; Euryarchaeota; Methanobacteria; Methanobacteriales;
OC Methanobacteriaceae; Methanothermobacter.
OX NCBI_TaxID=187420;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=Delta H;
RC MEDLINE=98037514; PubMed=9371463;
RX Smith D.R., Doucette-Stamm L.A., Deloughery C., Lee H.-M., Dubois J.,
RA Aldredge T., Bashirzadeh R., Blakely D., Cook R., Gilbert K.,
RA Harrison D., Hoang L., Keagle P., Lum W., Pothier B., Qiu D.,
RA Spadafora R., Vicare R., Wang Y., Wierzbowski J., Gibson R.,
RA Jiwaani N., Caruso A., Bush D., Safer H., Patwell D., Prabhakar S.,
RA McDougall S., Shmer G., Goyal A., Pietrovski S., Church G.M.,
RA Daniels C.J., Mao J.-I., Rice P., Noelling J., Reeve J.N.;
RT "Complete genome sequence of Methanobacterium thermoautotrophicum
RT deltaH: functional analysis and comparative genomics.";
RL J. Bacteriol. 179:7135-7155(1997).
CC -!- FUNCTION: CATALYZES THE REVERSIBLE OXIDATION OF CO(2) AND
CC METHANOFURAN (MFR) TO N-FORMYLMETHANOFURAN (CHO-MFR). CAN ONLY
CC OXIDISE FORMYLMETHANOFURAN. THIS ENZYME IS OXYGEN-LABILE.
CC -!- CATALYTIC ACTIVITY: Formylmethanofuran + H(2)O + acceptor = CO(2)
CC + methanofuran + reduced acceptor.
CC -!- COFACTOR: Tungsten.
CC -!- ENZYME REGULATION: Not inactivated by cyanide.
CC -!- PATHWAY: Methanogenesis; first step.
CC -!- SUBUNIT: THIS ENZYME IS COMPOSED OF SEVEN SUBUNITS FWDA (65 kDa),
CC FWDB (53 kDa), FWDC (31 kDa), FWDE (15 kDa), FWDF, AND FWDG.
CC -!- INDUCTION: By growth on tungsten or molybdenum under anaerobic
CC conditions.

```



FT SIGNAL 1 19 PROBABLE.  
FT CHAIN 20 378 CIRCUMSPOROZOITE PROTEIN.  
FT DOMAIN 99 211 18 X 6 AA TANDEM REPEATS OF D-G-A-R-A-  
[EA].  
FT REPEAT 99 104 1-1.  
FT REPEAT 105 110 1-2.  
FT REPEAT 111 116 1-3.  
FT REPEAT 117 122 1-4.  
FT REPEAT 123 128 1-5.  
FT REPEAT 129 134 1-6.  
FT REPEAT 135 140 1-7.  
FT REPEAT 141 146 1-8.  
FT REPEAT 147 152 1-9.  
FT REPEAT 153 158 1-10.  
FT REPEAT 159 164 1-11.  
FT REPEAT 165 170 1-12.  
FT REPEAT 171 176 1-13.  
FT REPEAT 177 182 1-14.  
FT REPEAT 183 188 1-15.  
FT REPEAT 189 194 1-16.  
FT REPEAT 195 200 1-17.  
FT REPEAT 201 206 1-18.  
FT DOMAIN 212 277 6 X 11 AA TANDEM REPEATS OF G-N-[QR]-  
[AE]-G-G-Q-A-G-A-G.  
FT REPEAT 212 222 2-1.  
FT REPEAT 223 233 2-2.  
FT REPEAT 234 244 2-3.  
FT REPEAT 245 255 2-4.  
FT REPEAT 256 266 2-5.  
FT REPEAT 267 277 2-6.  
FT DOMAIN 304 356 TSP TYPE-1.  
SQ SEQUENCE 378 AA; 37462 MW; 8295A913C36420C5 CRC64;

Query Match 69.6%; Score 39; DB 1; Length 378;  
Best Local Similarity 70.0%; Pred. No. 34;  
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 2 AGIGGEKAGG 11  
Db 263 AGAGGNRAGG 272

RESULT 9  
HXAA MOUSE STANDARD; PRT; 399 AA.  
AC P31310;  
DT 01-JUL-1993 (Rel. 26, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Homeobox protein Hox-A10 (Hox-1.8).  
GN HOXA10 OR HOXA-10 OR HOX-1.8.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxId=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX STRAIN=CD-1; TISSUE=Kidney;  
RX MEDLINE=95166244; PubMed=7862151;  
RA Benson G.V., Nguyen T.-H.E., Maas R.L.;  
RT "The expression pattern of the murine Hoxa-10 gene and the sequence  
recognition of its homeodomain reveal specific properties of  
Abdominal B-like genes";  
RL Mol. Cell. Biol. 15:1591-1601(1995).  
RN [2]  
RP SEQUENCE OF 325-384 FROM N.A.  
RX MEDLINE=92073356; PubMed=1683707;  
RA Singh G., Kaur S., Stock J.L., Jenkins N.A., Gilbert D.J.,  
RA Copeland N.G., Potter S.S.;  
RT "Identification of 10 murine homeobox genes";  
RL Proc. Natl. Acad. Sci. U.S.A. 88:10706-10710(1991).  
RN [3]  
RP SEQUENCE OF 346-370 FROM N.A.

RC STRAIN=C57BL/6; TISSUE=Spleen;  
RX MEDLINE=92073357; PubMed=1720547;  
RA Murtha M.T., Leckman J.F., Ruddle F.H.;  
RT "Detection of homeobox genes in development and evolution.";  
RL Proc. Natl. Acad. Sci. U.S.A. 88:10711-10715(1991).  
CC -!- FUNCTION: SEQUENCE-SPECIFIC TRANSCRIPTION FACTOR WHICH IS PART OF  
A DEVELOPMENTAL REGULATORY SYSTEM THAT PROVIDES CELLS WITH  
SPECIFIC POSITIONAL IDENTITIES ON THE ANTERIOR-POSTERIOR AXIS.  
CC BINDS TO THE DNA SEQUENCE 5'-AA[AT]TTTATTAC-3'.  
CC -!- SUBCELLULAR LOCATION: Nuclear.  
CC -!- ALTERNATIVE PRODUCTS:  
Event=Alternative splicing; Named isoforms=2;  
Name=1;  
IsoId=P31310-1; Sequence=Displayed;  
Name=2;  
IsoId=P31310-2; Sequence=VSP\_002386, VSP\_002387;  
CC -!- TISSUE SPECIFICITY: EXPRESSED IN THE DEVELOPING LIMB BUD WHERE IT  
IS RESTRICTED TO THE MESENCHYME ALONG THE PROXIMAL-DISTAL AXIS.  
CC ALSO FOUND IN DEVELOPING GUT AND UROGENITAL TRACT. IN ADULT  
TISSUE, BOTH FORMS FOUND IN KIDNEY BUT ONLY ISOFORM 1 IS EXPRESSED  
IN SKELETAL MUSCLE.  
CC -!- DEVELOPMENTAL STAGE: EMBRYONIC EXPRESSION INCREASES FROM DAY 9 TO  
DAY 12 AND THEN DECLINES TO DAY 15.  
CC -!- SIMILARITY: Belongs to the Abd-B homeobox family.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL; L08757; AAA67125.1; -.  
DR EMBL; M81659; AAA63312.1; -.  
DR PIR; I49754; I49754.  
DR HSP; P02833; 9ANT.  
DR TRANSFAC; T01715; -.  
DR MGD; MGI:96171; Hoxa10.  
DR InterPro; IPR001356; Homeobox.  
DR Pfam; PF00046; homeobox; 1.  
DR PRINTS; PR00024; HOMEBOX.  
DR ProDom; PD00010; Homeobox; 1.  
DR SMART; SM00389; HOX; 1.  
DR PROSITE; PS00027; HOMEBOX\_1; 1.  
DR PROSITE; PS00071; HOMEBOX\_2; 1.  
KW Homeobox; DNA-binding; Developmental protein; Nuclear protein;  
Transcription regulation; Alternative splicing.  
FT DOMAIN 26 38 GLY-RICH.  
FT DOMAIN 77 90 POLY-GLY.  
FT DOMAIN 113 144 GLN/PRO-RICH.  
FT DOMAIN 215 222 POLY-GLY.  
FT DOMAIN 271 274 POLY-GLY.  
FT DNA\_BIND 325 384 HOMEBOX.  
FT VARSPPLIC 1 305 Missing (in isoform 2).  
FT FTID=VSP\_002386.  
FT VARSPLIC 306 308 DSL-> MCQ (in isoform 2).  
FT FTID=VSP\_002387.  
SQ SEQUENCE 399 AA; 41415 MW; 7529624FC6057042 CRC64;

Query Match 69.6%; Score 39; DB 1; Length 399;  
Best Local Similarity 70.0%; Pred. No. 36;  
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 2 AGIGGEKAGG 11  
Db 25 AGVGGGSAGG 34

RESULT 10  
FUR2 DROME  
ID FUR2 DROME STANDARD; PRT; 1679 AA.  
AC P30432; Q24301;



FT CARBOHYD 1274 1274 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1277 1277 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1439 1439 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CONFLICT 152 153 MISSING (IN REF. 1).  
 FT CONFLICT 177 177 V -> F (IN REF. 1).  
 FT CONFLICT 213 213 V -> VOOL (IN REF. 1).  
 SQ SEQUENCE 1679 AA; 183369 MW; 3F9E749F0B021CF6 CRC64;

Query Match 69.6%; Score 39; DB 1; Length 1679;  
 Best Local Similarity 60.0%; Pred. No. 1.3e+02;

Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 1 IAGIGGKAG 10  
 Db 176 LVGLGGERAG 185

## RESULT 11

TXB2\_CABEL ID TXB2\_CABEL STANDARD; PRT; 423 AA.  
 AC Q19691;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Putative T-box protein 2.  
 GN TXB-2 OR F21H11.3  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoides;  
 OC Rhabditidae; Peloderinae; Caenorhabditis.  
 OX NCBI\_TaxID=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Bristol N2;  
 RA Tang H.;  
 RT Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.  
 RL Nature 346:155-156 (1992).  
 CC -!- SUBCELLULAR LOCATION: Nuclear (Potential).  
 CC -!- SIMILARITY: Contains 1 T-box domain.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

EMBL; U11279; AAF99909.1; -.  
 PIR; B88450; B88450.  
 DR HSSP; P24781; 1XBR.  
 DR TRANSFAC; T04406; -.  
 DR WormPep; F21H11.3; CE01245.  
 DR InterPro; IPR008967; P53-like.  
 DR InterPro; IPR001699; TF T-box.  
 DR Pfam; PF00907; T-box; 1.  
 DR PRINTS; PR00937; TBOX.  
 DR SMART; SM00425; TBOX; 1.  
 DR PROSITE; PS01283; TBOX\_1; 1.  
 DR PROSITE; PS01264; TBOX\_2; 1.  
 DR PROSITE; PS0252; TBOX\_3; 1.  
 DR PROSITE; PS0252; TBOX\_3; 1.  
 KW Hypothetical protein; DNA-binding; Nuclear protein.  
 FT DNA\_BIND 70 243 T-BOX.  
 FT DOMAIN 365 368 POLY-ALA.  
 FT SEQUENCE 423 AA; 46997 MW; 91FD17596F211F40 CRC64;

Query Match 67.9%; Score 38; DB 1; Length 423;  
 Best Local Similarity 70.0%; Pred. No. 55;  
 Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 2 AGIGGKAGG 11  
 Db 20 AGVGGGAGG 29

RESULT 12  
 AMPA\_PSEAE ID AMPA\_PSEAE STANDARD; PRT; 495 AA.  
 AC O68822;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Cytosol aminopeptidase (EC 3.4.11.1) (Leucine aminopeptidase) (LAP)  
 DE (Leucyl aminopeptidase).  
 GN PEPA OR PHPA OR PA3931.  
 OS Pseudomonas aeruginosa.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
 OC Pseudomonadaceae; Pseudomonas.  
 OX NCBI\_TaxID=287;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=FRD1;  
 RX MEDLINE=99084946; PubMed=9864319;  
 RA Woolwine S.C.; Wozniak D.J.;  
 RT "Identification of an Escherichia coli pepA homolog and its  
 RT involvement in suppression of the algs phenotype in mucoid Pseudomonas  
 RT aeruginosa.";  
 RL J. Bacteriol. 181:107-116(1999).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 15692 / PA01;  
 RX MEDLINE=20437337; PubMed=10984043;  
 RA Stover C.K.; Pham X.-Q.T.; Erwin A.L.; Mizoguchi S.D.; Warren P.;  
 RA Hickey M.J.; Brinkman F.S.L.; Hufnagle W.O.; Kowalik D.J.; Lagrou M.;  
 RA Garber R.L.; Goltry L.; Tolentino E.; Westbrook-Wadman S.; Yuan Y.;  
 RA Brody L.L.; Coulter S.N.; Folger K.R.; Kas A.; Larbig K.; Lim R.M.;  
 RA Smith K.A.; Spencer D.H.; Wong G.K.-S.; Wu Z.; Paulsen I.T.;  
 RA Reizer J.; Sailer M.H.; Hancock R.E.W.; Lory S.; Olson M.V.;  
 RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an  
 RT opportunistic pathogen.";  
 RL Nature 406:959-964(2000).  
 CC -!- FUNCTION: Presumably involved in the processing and regular  
 CC turnover of intracellular proteins. Catalyzes the removal of  
 CC unsubstituted N-terminal amino acids from various peptides (By  
 CC similarity).  
 CC -!- CATALYTIC ACTIVITY: Release of an N-terminal amino acid, Xaa-|-  
 CC Xbb-, in which Xaa is preferably Leu, but may be other amino acids  
 CC including Pro although not Arg or Lys, and Xbb may be Pro.  
 CC -!- COFACTOR: Binds 2 manganese ions per subunit (By similarity).  
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).  
 CC -!- SIMILARITY: Belongs to peptidase family M17.  
 CC -----

This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

EMBL; AF054622; AAD04821.1; -.  
 DR EMBL; AE004800; AAG07218.1; -.  
 DR PIR; D83167; D83167.  
 DR HSSP; P00727; 1LAM.  
 DR MEMOPS; M17.003; -.  
 DR HAMAP; MF 00181; -; 1.  
 DR InterPro; IPR000819; Peptidase M17 C.  
 DR InterPro; IPR008283; Peptidase M17\_N.  
 DR Pfam; PF00883; Peptidase M17; 1.  
 DR Pfam; PF02789; Peptidase M17\_N; 1.  
 DR PRINTS; PR00481; LAMNOPPTDASE.  
 DR PROSITE; PS00631; CYTOSOL\_AP; 1.  
 DR Hydrolase; Aminopeptidase; Manganese; Complete proteome.  
 KW METAL 266 266 MANGANESE 2 (BY SIMILARITY).  
 FT METAL 271 271 MANGANESE 1 AND 2 (BY SIMILARITY).  
 FT METAL 289 289 MANGANESE 2 (BY SIMILARITY).  
 FT METAL 348 348 MANGANESE 1 (BY SIMILARITY).  
 FT METAL 350 350 MANGANESE 1 AND 2 (BY SIMILARITY).



FT ACT SITE 278 278 POTENTIAL.  
 FT ACT SITE 352 352 POTENTIAL.  
 SQ SEQUENCE 495 AA; 52331 MW; E08C7928E7A0FDCB CRC64;  
 Query Match 67.9%; Score 38; DB 1; Length 495;  
 Best Local Similarity 80.0%; Pred. No. 64;  
 Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 IAGIGGEKAG 10  
 Db 433 IANIGGPKAG 442  
 RESULT 13  
 POLG PWMI STANDARD; PRT; 269 AA.  
 AC P32574;  
 DT 01-OCT-1993 (Rel. 27, Last sequence update)  
 DT 01-OCT-1993 (Rel. 27, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Genome polyprotein [Contains: Coat protein (CP)] (Fragment).  
 OS Passionfruit woodiness virus (strain Mild) (PWV).  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Potyviridae;  
 OC Potyvirus.  
 OX NCBI\_TaxID=31733;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=89075954; PubMed=3202695;  
 RA Shukla D.D., McKern N.M., Ward C.W.;  
 RT "Coat protein of potyviruses. 5. Symptomatology, serology, and coat  
 protein sequences of three strains of passionfruit woodiness virus.";  
 RL Arch. Virol. 102:221-232(1988).  
 CC -1- PPM: The viral RNA of potyviruses is expressed as a single  
 CC polyprotein which undergoes posttranslational proteolytic  
 CC processing resulting in the production of at least eight  
 CC individual proteins.  
 CC -1- SIMILARITY: Belongs to the potyviruses polyprotein family.  
 DR PIR; C60078; VCVSWM.  
 DR InterPro; IPR001592; Poty\_coat.  
 DR Pfam; PF00767; Poty\_coat; 1.  
 KW Coat protein; Polyprotein.  
 FT NON\_TER 1 1  
 SQ SEQUENCE 269 AA; 29934 MW; B2B4E542267FA919 CRC64;  
 Query Match 66.1%; Score 37; DB 1; Length 269;  
 Best Local Similarity 70.0%; Pred. No. 53;  
 Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
 QY 2 AGIGGEKAGG 11  
 Db 7 AGIDGKGGG 16  
 RESULT 14  
 RFI LEPIN STANDARD; PRT; 354 AA.  
 AC 08F956;  
 DT 10-OCT-2003 (Rel. 42, Created)  
 DT 10-OCT-2003 (Rel. 42, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Peptide chain release factor 1 (RP-1).  
 GN PFRA OR LA0114.  
 OS Leptospira interrogans.  
 OC Bacteria; Spirochaetes; Spirochaetales; Leptospiraceae; Leptospira.  
 OX NCBI\_TaxID=173;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=56601 / Serogroup Icterohaemorrhagiae / Serovar lai;  
 RX MEDLINE=22598143; PubMed=12712204;  
 RA Ren S.-X., Fu G., Jiang X.-G., Zeng R., Miao Y.-G., Xu H.,  
 RA Zhang Y.-X., Xiong H., Lu G., Lu L.-F., Jiang H.-Q., Jia J., Tu Y.-F.,  
 RA Jiang J.-X., Gu W.-Y., Zhang Y.-Q., Cai Z., Sheng H.-H., Yin H.-F.,  
 RA Zhang Y., Zhu G.-F., Wan M., Huang H.-L., Qian Z., Wang S.-Y., Ma W.,

RA Yao Z.-J., Shen Y., Qiang B.-Q., Xia Q.-C., Guo X.-K., Danchin A.,  
 RA Saint Girons I., Somerville R.L., Wen Y.-M., Shi M.-H., Chen Z.,  
 RA Xu J.-G., Zhao G.-P.;  
 RT "Unique Physiological and pathogenic features of Leptospira  
 RT interorgans revealed by whole-genome sequencing.";  
 RL Nature 422:888-893(2003).  
 CC -1- FUNCTION: Peptide chain release factor 1 directs the termination  
 CC of translation in response to the peptide chain termination codons  
 CC UAG and UAA (By similarity).  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.  
 CC -1- SIMILARITY: Belongs to the prokaryotic/mitochondrial release  
 CC factor family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AB011202; AAM47313.1; -.  
 DR HAMAP; MF 00093; -; 1.  
 DR InterPro; IPR005139; PCRF.  
 DR InterPro; IPR000352; Pep\_rel\_factor\_I.  
 DR InterPro; IPR004373; PrfA.  
 DR Pfam; PF03462; PCRF; 1.  
 DR Pfam; PF00472; RF-1; 1.  
 DR TIGRfam; TIGR00019; prfA; 1.  
 DR PROSITE; PS00745; RF\_PROK\_I; 1.  
 KW Protein biosynthesis; Complete proteome.  
 SQ SEQUENCE 354 AA; 40099 MW; 779A91ACCC997A59 CRC64;  
 Query Match 66.1%; Score 37; DB 1; Length 354;  
 Best Local Similarity 77.8%; Pred. No. 68;  
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
 QY 2 AGIGGEKAG 10  
 Db 113 AGTGGREAG 121  
 RESULT 15  
 TRMU YERPE STANDARD; PRT; 371 AA.  
 AC Q8ZPQ3;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Probable tRNA (5-methylaminomethyl-2-thiouridyate)-methyltransferase  
 DE (EC 2.1.1.61).  
 GN TRMU OR YPO1638 OR Y1799.  
 OS Yersinia pestis.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Yersinia.  
 OX NCBI\_TaxID=632;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CO-92 / Biovar Orientalis;  
 RX MEDLINE=21470413; PubMed=11586360;  
 RA Parkhill J., Wren B.W., Thomson N.R., Titball R.W., Holden M.T.G.,  
 RA Prentice M.B., Sebahia M., James K.D., Churcher C., Mungall K.L.,  
 RA Baker S., Basham D., Bentley S.D., Brooks K., Cerdano-Tarraga A.M.,  
 RA Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,  
 RA Felwell T., Hamlin N., Holroyd S., Jagsels K., Karlyshev A.V.,  
 RA Leather S., Moule S., Oyston P.C.F., Quail M.A., Rutherford K.,  
 RA Simmonds M., Skelton J., Stevens K., Whitehead S., Barrell B.G.;  
 RT "Genome sequence of Yersinia pestis, the causative agent of plague.";  
 RL Nature 413:523-527(2001).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=KIMS / Biovar Mediaevalis;  
 RX MEDLINE=22137863; PubMed=12142430;



RA Deng W., Burland V., Plunkett G. III, Boutin A., Mayhew G.F., Liss P.,  
 RA Perna N.T., Rose D.J., Mau B., Zhou S., Schwartz D.C.,  
 RA Fetherston J.D., Lindler L.E., Brubaker R.R., Plano G.V.,  
 RA Straley S.C., McDonough K.A., Nilles M.L., Matson J.S., Blattner F.R.,  
 RA Perry R.D.;  
 RA "Genome sequence of Yersinia pestis KIM.";  
 RL J. Bacteriol. 184:4601-4611(2002).  
 CC -!- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + tRNA = S-adenosyl-L-  
 CC homocysteine + tRNA containing 5-methylaminomethyl-2-  
 CC thiouridylate.  
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).  
 CC -!- SIMILARITY: Belongs to the trmU family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: AJ414149; CAC90460.1; -;  
 DR EMBL: AE013782; AAM85367.1; -;  
 DR PIR: A10199; A10199..  
 DR HAWAP; MF\_00144; -; 1.  
 DR InterPro; IPR004506; TrmU.  
 DR Pfam; PF03054; tRNA\_Me\_trans; 1.  
 DR TIGRFAMs; TIGR00420; trmU; 1.  
 KW Transferase; Methyltransferase; tRNA processing; Complete proteome.  
 SQ SEQUENCE 371 AA; 41365 MW; D7A301D70B2E4F5F CRC64;

Query Match 66.1%; Score 37; DB 1; Length 371;  
 Best Local Similarity 77.8%; Pred. NO. 71;  
 Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 3 GIGGEKAGG 11  
 |||||  
 Db 248 GIGGTKEGG 256

Search completed: September 18, 2004, 04:26:24  
 Job time : 10.1961 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 18, 2004, 04:00:08 ; Search time 39.0392 Seconds  
(without alignments)  
88.903 Million cell updates/sec

Title: US-10-615-959-42  
Perfect score: 56  
Sequence: 1 IAGIGSEKAGG 11

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SPTREMBL 25;\*

1: sp\_archea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_nhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_virus:\*  
16: sp\_bacteriap:\*  
17: sp\_archeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	56	100.0	1163	Q8N6U4	Q8N6U4 homo sapien
2	52	92.9	338	11 Q8BJU6	Q8BJU6 mus musculus
3	52	92.9	1222	11 Q8K173	Q8K173 mus musculus
4	52	92.9	1464	11 Q8BKV2	Q8BKV2 mus musculus
5	52	92.9	1464	11 Q8BLW4	Q8BLW4 mus musculus
6	48	85.7	1464	11 Q7TT32	Q7TT32 mus musculus
7	44	78.6	487	16 Q53174	Q53174 mycobacteri
8	44	78.6	487	16 Q7TYK1	Q7TYK1 mycobacteri
9	42	75.0	209	16 Q8UAC0	Q8UAC0 agrobacteri
10	41	73.2	208	10 Q8L418	Q8L418 oryza sativ
11	41	73.2	460	2 Q8RS19	Q8RS19 gamma-prote
12	40	71.4	136	11 Q9D3C5	Q9D3C5 mus musculus
13	40	71.4	156	10 Q9FRJ4	Q9FRJ4 oryza sativ
14	40	71.4	156	10 Q7XCM3	Q7XCM3 oryza sativ
15	40	71.4	298	16 Q930J8	Q930J8 rhizobium m
16	40	71.4	548	3 Q9P3K6	Q9P3K6 neurospora

## RESULT 1

Q8N6U4  
ID Q8N6U4 PRELIMINARY; PRT; 1163 AA.  
AC Q8N6U4;  
DT 01-OCT-2002 (TReMBLrel. 22, Created)  
DT 01-OCT-2002 (TReMBLrel. 22, Last sequence update)  
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)  
DE Similar to collagen, type III, alpha 1 (Ehlers-Danlos syndrome type IV, autosomal dominant).  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Liver;  
RA Strausberg R.;  
RL Submitted (APR-2002) to the EMBL/GenBank/DBSJ databases.  
DR EMBL; BC028178; AAH28178.1; -  
DR GO; GO:0005581; C:collagen; IEA.  
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.  
DR InterPro; IPR000160; Collagen.  
DR InterPro; IPR002181; Fibrinogen\_C.  
DR InterPro; IPR001007; VWF\_C.  
DR Pfam; PF01410; COLFI; 1.  
DR Pfam; PF01391; Collagen; 13.  
DR ProDom; PD002078; Fib\_collagen\_C; 1.  
DR SMART; SM00038; COLFI; 1.  
DR SMART; SM00214; VWC; 1.  
DR PROSITE; PS01208; VWF\_C\_1; 1.  
DR PROSITE; PS0184; VWF\_C\_2; 1.  
KW Collagen.  
SQ SEQUENCE 1163 AA; 111899 MW; 9E0C6BE1E94D6357 CRC64;

Query Match 100.0%; Score 56; DB 4; Length 1163;

Best Local Similarity 100.0%; Pred. No. 1.4;

Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

## ALIGNMENTS

```
QY 1 IAGIGGEKAGG 11
Db 902 IAGIGGEKAGG 912

RESULT 2
Q8BUJ6 PRELIMINARY; PRT; 338 AA.
ID Q8BUJ6
AC Q8BUJ6; TISSUE=Embryo;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Collagen alpha 1 (Fragment).
GN COL3A1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Embryo;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs."
RL Nature 420:563-573 (2002).
DR EMBL; AK079113; BAC37545.1; -.
DR MGD; MGI:88453; Col3a1.
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR002181; Fibrinogen_C.
DR Pfam; PF01410; COLFI; 1.
DR ProDom; PD002078; Fib.collagen_C; 1.
DR SMART; SM00038; COLFI; 1.
FT NON_TER 1
SQ SEQUENCE 338 AA; 35813 MW; 15B8369D8690F37A CRC64;

Query Match 92.9%; Score 52; DB 11; Length 338;
Best Local Similarity 81.8%; Pred. No. 1.7;
Matches 9; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 IAGIGGEKAGG 11
Db 77 IAGVGGEKSGG 87

RESULT 3
Q8KL73 PRELIMINARY; PRT; 1222 AA.
AC Q8KL73
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein (Fragment).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Breast tumor;
RA Strausberg R.;
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC028248; AAR28248.1; -.
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR InterPro; IPR008161; Clg_helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR002181; Fibrinogen_C.
```

```
DR InterPro; IPR000885; Fib_collagen_C.
DR Pfam; PF01410; COLFI; 1.
DR Pfam; PF01391; Collagen; 16.
DR ProDom; PD000007; Clg_helix; 1.
DR ProDom; PD002078; Fib_collagen_C; 1.
DR SMART; SM00038; COLFI; 1.
KW Hypothetical protein; Collagen.
FT NON_TER 1
SQ SEQUENCE 1222 AA; 115140 MW; A409CA00D82765E4 CRC64;

Query Match 92.9%; Score 52; DB 11; Length 1222;
Best Local Similarity 81.8%; Pred. No. 6.6;
Matches 9; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 IAGIGGEKAGG 11
Db 961 IAGVGGEKSGG 971

RESULT 4
Q8BKY2 PRELIMINARY; PRT; 1464 AA.
AC Q8BKY2
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Collagen alpha 1.
GN COL3A1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Head;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs."
RL Nature 420:563-573 (2002).
DR EMBL; AK048546; BAC33370.1; -.
DR MGD; MGI:88453; Col3a1.
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR InterPro; IPR008161; Clg_helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR002181; Fibrinogen_C.
DR InterPro; IPR000885; Fib_collagen_C.
DR InterPro; IPR001007; VWF_C.
DR Pfam; PF01410; COLFI; 1.
DR Pfam; PF01391; Collagen; 18.
DR ProDom; PD000007; Clg_helix; 1.
DR ProDom; PD002078; Fib_collagen_C; 1.
DR SMART; SM00038; COLFI; 1.
DR SMART; SM00214; VWC; 1.
DR PROSITE; PS01208; VWFC_1; 1.
DR PROSITE; PS0184; VWFC_2; 1.
SQ SEQUENCE 1464 AA; 138947 MW; 1E4ED9539EF42C12 CRC64;

Query Match 92.9%; Score 52; DB 11; Length 1464;
Best Local Similarity 81.8%; Pred. No. 8;
Matches 9; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 IAGIGGEKAGG 11
Db 1203 IAGVGGEKSGG 1213

RESULT 5
Q8BLM4 PRELIMINARY; PRT; 1464 AA.
AC Q8BLM4
```

DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Collagen alpha 1.  
 GN COL3A1.  
 OS Mus musculus (Mouse)  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Aorta, and Vein;  
 RX MEDLINE=22354683; PubMed=12466851;  
 RA The FANTOM Consortium,  
 RA "Analysis of the mouse transcriptome based on functional annotation of  
 RT 60,770 full-length cDNAs."  
 RL Nature 420:563-573 (2002).  
 DR EMBL; AK041115; BAC30826.1; --  
 DR MGI; MGI:88453; Col3a1.  
 DR GO; GO:0005581; C:collagen; IEA.  
 DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.  
 DR InterPro; IPR008161; Clg\_helix.  
 DR InterPro; IPR008160; Collagen.  
 DR InterPro; IPR002181; Fibrinogen\_C.  
 DR InterPro; IPR000885; Fib\_collagen\_C.  
 DR InterPro; IPR001007; VWF\_C.  
 DR Pfam; PF01410; COLFI; 1.  
 DR Pfam; PF01391; Collagen; 18.  
 DR ProDom; PD000007; Clg\_helix; 1.  
 DR SMART; SM002078; Fib\_collagen\_C; 1.  
 DR SMART; SM00038; COLFI; 1.  
 DR SMART; SM00214; VMC; 1.  
 DR PROSITE; PS01208; VWF\_C; 1.  
 DR PROSITE; PS01084; VWF\_C; 1.  
 SQ SEQUENCE 1464 AA; 138939 MW; 91F3246D90818449 CRC64;  
  
 Query Match 92.9%; Score 52; DB 11; Length 1464;  
 Best Local Similarity 81.8%; Pred. No. 8;  
 Matches 9; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
  
 OY 1 IAGIGGKAGG 11  
 DB 1203 IAGVGGEKSGG 1213  
  
 RESULT 6  
 DT 01-OCT-2003 (TrEMBLrel. 25, Created)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Hypothetical protein.  
 OS Mus musculus (Mouse)  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Brain;  
 RX MEDLINE=22388257; PubMed=12477932;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Frange C.,  
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,

RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,  
 RA Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,  
 RA Jones S.J., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length human  
 and mouse cDNA sequences."  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Brain;  
 RA Strausberg R.;  
 RL Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; BC052398; AAHS2398.1; --  
 KW Hypothetical protein.  
 SQ SEQUENCE 1464 AA; 138971 MW; 2B38BC27AF21590B CRC64;  
  
 Query Match 85.7%; Score 48; DB 11; Length 1464;  
 Best Local Similarity 72.7%; Pred. No. 37;  
 Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
  
 OY 1 IAGIGGKAGG 11  
 DB 1203 IAGVGGEKSGG 1213  
  
 RESULT 7  
 DT 01-JUN-1998 (TrEMBLrel. 06, Created)  
 DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Foliopolylglutamate synthase.  
 GN POLC OR RV2447C OR MTV008.03C OR MT2523.  
 OS Mycobacterium tuberculosis.  
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
 OX NCBI\_TaxID=1773;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=H37Rv;  
 RX MEDLINE=98295987; PubMed=9634230;  
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
 RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,  
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,  
 RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,  
 RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,  
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
 RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,  
 RA Sulston J.E., Taylor K., Whitehead S., Barrett B.G.;  
 RT "Deciphering the biology of Mycobacterium tuberculosis from the  
 complete genome sequence."  
 RL Nature 393:537-544 (1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CDC 1551 / Oshkosh;  
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,  
 RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,  
 RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,  
 RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,  
 RA Bishai W.;  
 RT "Whole genome comparison of Mycobacterium tuberculosis clinical and  
 laboratory strains."  
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AL021245; CAAL6024.1; --  
 DR EMBL; AE007089; AAK46822.1; --  
 DR FIR; F70863; F70863.  
 DR HSSP; P15925; 1FGS.  
 DR TIGR; MT2523; --  
 DR TubercuList; Rv2447c; --  
 DR GO; GO:0005524; F:ATP binding; IEA.

```

DR GO: GO:0016874; F:ligase activity; IEA.
DR GO: GO:0004326; F:tetrahydrofolylpolyglutamate synthase activity; IEA.
DR GO: GO:0009058; P:biosynthesis; IEA.
DR GO: GO:0009396; P:folic acid and derivative biosynthesis; IEA.
DR InterPro: IPR001645; F:polygl synthase.
DR InterPro: IPR000713; Mur_ligase.
DR InterPro: IPR004101; Mur_ligase_C.
DR Pfam: PF01225; Mur_ligase_1.
DR Pfam: PF02875; Mur_ligase_C; 1.
DR TIGRFAMs: TIGR01495; folC; 1.
DR PROSITE: PS01012; FOLYLPOLYGLU_SYNT_2; 1.
KW Complete proteome.
SQ SEQUENCE 487 AA; 50779 MW; 77D3A4E3D2DC59D0 CRC64;

Query Match 78.6%; Score 44; DB 16; Length 487;
Best Local Similarity 90.0%; Pred. No. 53;
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 IAGIGGEKAG 10
Db 209 IAGIAGEKAG 218

RESULT 8
Q7TYK1 PRELIMINARY; PRT; 487 AA.
ID Q7TYK1 PRELIMINARY; PRT; 487 AA.
AC Q7TYK1
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Probable folylpolyglutamate synthase protein FOLC
DE (Poly)poly-gamma-glutamate synthetase (FPGS) (EC 6.3.2.17).
GN FOLC OR MB2474C.
OS Mycobacterium bovis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1765;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AF2122/97;
RX MEDLINE=22709107; PubMed=12788972;
RA Garnier T., Eiglmeyer K., Camus J.-C., Medina N., Mansoor H.,
RA Pryor M., Duthoy S., Grondin S., Lacroix C., Monsemp C., Simon S.,
RA Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,
RA Parkhill J., Barrall B.G., Cole S.T., Gordon S.V., Hewinson R.G.;
RT "The complete genome sequence of Mycobacterium bovis.";
RL Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882(2003).
DR EMBL; BX248342; CAD97335.1; -.
KW Ligase; Complete proteome.
SQ SEQUENCE 487 AA; 50779 MW; 77D3A4E3D2DC59D0 CRC64;

Query Match 78.6%; Score 44; DB 16; Length 487;
Best Local Similarity 90.0%; Pred. No. 53;
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 IAGIGGEKAG 10
Db 209 IAGIAGEKAG 218

RESULT 9
Q8UAC0 PRELIMINARY; PRT; 209 AA.
ID Q8UAC0 PRELIMINARY; PRT; 209 AA.
AC Q8UAC0;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Arylesterase.
GN ADA OR ATU3454 OR AGR_L2749.
OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Rhizobium/Agrobacterium group; Agrobacterium.
OX NCBI_TaxID=176299;

```

```

RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21608550; PubMed=11743193;
RA Wood D.W., Setubal J.C., Kaul R., Monks D.E., Kitajima J.P.,
RA Okura V.K., Zhou Y., Chen L., Wood G.E., Almeida N.F., Jr., Woo L.,
RA Chen Y., Paulsen I.T., Eisen J.A., Karp P.D., Bovee D. Sr.,
RA Chapman P., Clendenning J., Deatherage G., Gillet W., Grant C.,
RA Kutayav T., Levy R., Li M.-J., Mclelland E., Palmieri A.,
RA Raymond C., Rouse G., Saenphimmachak C., Wu Z., Romero P., Gordon D.,
RA Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Perry M.,
RA Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,
RA Chumley P., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,
RA Nester E.W.;
RT "The genome of the natural genetic engineer Agrobacterium tumefaciens
RT C58.";
RL Science 294:2317-2323(2001).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=21608551; PubMed=11743194;
RA Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M.,
RA Quorlo B., Goldman B.S., Cao Y., Askenazi M., Tatchouk O., Epp A., Liu F.,
RA Houmlel K., Gordon J., Vaudin M., Tatchouk O., Epp A., Liu F.,
RA Wollam C., Allinger M., Doughty D., Scott C., Lappas C., Markelz B.,
RA Flanagan C., Crowell C., Gurson J., Lomo C., Sear C., Strub G.,
RA Cielo C., Slater S.;
RT "Genome sequence of the plant pathogen and biotechnology agent
RT Agrobacterium tumefaciens C58.";
RL Science 294:2323-2328(2001).
DR EMBL; AE009275; AAL44267.1; -.
DR EMBL; AE008336; AAK89941.1; -.
DR PIR; AE2981; AE2981.
DR PIR; C98302; C98302.
DR GO: GO:0003824; F:catalytic activity; IEA.
DR InterPro: IPR001087; Lipase_GDSL.
DR Pfam; PF00657; Lipase_GDSL; 1.
KW Complete proteome.
SQ SEQUENCE 209 AA; 21865 MW; B09845588049D0C6 CRC64;

Query Match 75.0%; Score 42; DB 16; Length 209;
Best Local Similarity 72.7%; Pred. No. 47;
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 IAGIGGEKAG 11
Db 137 VAGPGGEPAGG 147

RESULT 10
Q8L418 PRELIMINARY; PRT; 208 AA.
ID Q8L418 PRELIMINARY; PRT; 208 AA.
AC Q8L418;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE P0551A11.6 protein (QJ1116 C07.6 protein).
GN P0551A11.6 OR QJ1116 C07.6_6
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Nipponbare;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 1, PAC
RT clone: P0551A11.";
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Nipponbare;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 1, BAC

```

RT clone:OJ1116 C07.";  
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AP003934; BAB92814.1; -;  
 DR EMBL; AP004253; BAC00689.1; -;  
 DR Gramene; Q8L418; -;  
 SQ SEQUENCE 208 AA; 20670 MW; C4D641DBA4D55B80 CRC64;

Query Match 73.2%; Score 41; DB 10; Length 208;  
 Best Local Similarity 54.5%; Pred. No. 68;  
 Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Qy 1 IAGIGGEKAGG 11  
 : : : : :  
 Db 63 VGGVGGREAGG 73

RESULT 11

Q8RSI9 PRELIMINARY; PRT; 460 AA.

AC Q8RSI9;  
 DT 01-JUN-2002 (TrEMBLrel. 21, Created)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE TrbI protein.  
 GN TRBI.  
 OS Gamma-proteobacterium Hot 75m4.  
 OG Plasmid pB4.  
 OC Bacteria; environmental samples.  
 OC NCBI\_TaxID=77133;  
 RN [1]  
 RP SEQUENCE FROM N.A.

RA Tauch A.;  
 RT "The IncP beta plasmid pB4 encodes a tripartite antibiotic efflux  
 RT system of the RND-MFP-OMP type conferring erythromycin and  
 RT roxithromycin resistance in *Pseudomonas* sp. B13."  
 RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.

DR EMBL; AJ341260; CA224375.1; -;  
 DR GO; GO:0046821; C:extrachromosomal DNA; IEA.  
 DR GO; GO:0009291; P:unidirectional conjugation; IEA.  
 DR InterPro; IPR005498; TrbI.  
 DR Pfam; PF03743; TrbI; 1.

KW Plasmid.  
 SQ SEQUENCE 460 AA; 47889 MW; B88772672063B91D CRC64;

Query Match 73.2%; Score 41; DB 2; Length 460;  
 Best Local Similarity 60.0%; Pred. No. 1.6e+02;  
 Matches 6; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 2 AGIGGEKAGG 11  
 : : : : :  
 Db 211 SGVGGQQAGG 220

RESULT 12

Q9D3C5 PRELIMINARY; PRT; 136 AA.

AC Q9D3C5;  
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
 DE 5830493J20Rik protein.  
 GN 5830493J20Rik.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OC NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.

RC STRAIN=C57BL/6J; TISSUE=Thymus;  
 RX MEDLINE=21085660; PubMed=11217851;  
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
 RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,  
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,

RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,  
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,  
 RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,  
 RA Schriml L.M., Staubli P., Suzuki R., Tonita M., Wagner L., Washio T.,  
 RA Sakai K., Okido T., Furuno M., Rono H., Baldarelli R., Barsh G.,  
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaudo M.F.,  
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,  
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,  
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
 RA Sasaki H., Sato K., Schoenbach C., Seva T., Shibata Y., Storch K.-P.,  
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,  
 RA Wyshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,  
 RA Hayashizaki Y.;  
 RT "Functional annotation of a full-length mouse cDNA collection.";  
 RL Nature 409:685-690(2001).  
 DR EMBL; AK018067; BAB31058.1; -;  
 DR MGD; MGI:1923370; 5830493J20Rik.

SQ SEQUENCE 136 AA; 13746 MW; B125F747DA12CFF2 CRC64;

Query Match 71.4%; Score 40; DB 11; Length 136;  
 Best Local Similarity 70.0%; Pred. No. 64;  
 Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 2 AGIGGEKAGG 11  
 : : : : :  
 Db 28 AGLGGRSAGG 37

RESULT 13

ID Q9FRJ4 PRELIMINARY; PRT; 156 AA.

AC Q9FRJ4;  
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Hypothetical protein.  
 GN OSJNB0064P21.10.  
 OS Oryza sativa (Rice).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC Ehrhartoideae; Oryzaceae; Oryza.  
 OC NCBI\_TaxID=4530;

RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Nipponbare;  
 RA Buell C.R., Yuan Q., Moffat K.S., Hill J.N., Burr P.C., Hsiao J.,  
 RA Zismann V., Pai G., Bowman C.L., Fujii C.Y., VanAken S.E.,  
 RA Bowman C.L., Craven B., Utterback T.R., Khalak H., Feldblyum T.V.,  
 RA Quackenbush J., White O., Salzberg S.L., Fraser C.M.;  
 RT "Oryza sativa chromosome 10 BAC OSJNB0064P21 genomic sequence."  
 RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL; AC073166; AAG46112.1; -;  
 DR Gramene; Q9FRJ4; -;  
 KW Hypothetical protein.

SQ SEQUENCE 156 AA; 15960 MW; 22DDB710603858FE CRC64;

Query Match 71.4%; Score 40; DB 10; Length 156;  
 Best Local Similarity 45.5%; Pred. No. 74;  
 Matches 5; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Qy 1 IAGIGGEKAGG 11  
 : : : : :  
 Db 114 VSGVGGSRSGG 124

RESULT 14

Q7XCM3

ID Q7XCM3 PRELIMINARY; PRT; 156 AA.

AC Q7XCM3;  
 DT 01-OCT-2003 (TrEMBLrel. 25, Created)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)

Search completed: September 18, 2004, 04:29:41  
Job time : 41.0392 secs

```

DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
GN OSUNBB0064P21.10.
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Nipponbare;
RA The Rice Chromosome 10 Sequencing Consortium;
RT "In-depth view of structure, activity, and evolution of rice
RL chromosome 10.";
RL Science 300:1566-1569(2003).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Nipponbare;
RA Buell C.R., Wing R.A., McCombie W.R., Messing J., Yuan Q.;
RL Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF017116; AAP54848.1; -.
KW Hypothetical protein.
SQ SEQUENCE 156 AA; 15960 MW; 22DDB710603858FE CRC64;

Query Match 71.4%; Score 40; DB 10; Length 156;
Best Local Similarity 45.5%; Pred. No. 74;
Matches 5; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

QY 1 IAGIGGEKAGG 11
Db 114 VSGVGGRSGG 124

RESULT 15
Q930J8 PRELIMINARY; PRT; 298 AA.
AC Q930J8;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Hypothetical protein RA0197.
GN RA0197 OR SMA0367.
OS Rhizobium meliloti (Sinorhizobium meliloti).
OG Plasmid pSymA (megaplasmid 1).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Sinorhizobium/Ensifer group; Sinorhizobium.
OX NCBI_TaxID=382;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1021;
RX MEDLINE=21396509; PubMed=11481432;
RA Barnett M.J., Fisher R.F., Jones T., Komp C., Abola A.P.,
RA Barloy-Hubler F., Bowser L., Capela D., Galibert F., Gouzy J.,
RA Gurjal M., Hong A., Huizar L., Hyman R.W., Kahn D., Kahn M.L.,
RA Kalman S., Keating D.H., Palm C., Peck M.C., Surzycki R., Wells D.H.,
RA Yeh K.-C., Davis R.W., Federspiel N.A., Long S.R.;
RT "Nucleotide sequence and predicted functions of the entire
RT Sinorhizobium meliloti pSymA megaplasmid.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:9883-9888(2001).
DR EMBL; AF007212; AAK64855.1; -.
DR PIR; E95286; E95286.
DR GO; GO:0046821; C:extrachromosomal DNA; IEA.
KW Hypothetical protein; Plasmid; Complete proteome.
SQ SEQUENCE 298 AA; 28617 MW; 27C9E966FD7D2068 CRC64;

Query Match 71.4%; Score 40; DB 16; Length 298;
Best Local Similarity 88.9%; Pred. No. 1.5e+02;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AGIGGEKAG 10
Db 103 AGIGGAKAG 111
```



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 18, 2004, 03:47:39 ; Search time 64.2353 Seconds  
(without alignments)  
52.784 Million cell updates/sec

Title: US-10-615-959-43

Perfect score: 62

Sequence: 1 IAGIGGEKAGGF 12

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_29Jan04.\*

1: Geneseq1980s.\*

2: Geneseq1990s.\*

3: Geneseq2000s.\*

4: Geneseq2001s.\*

5: Geneseq2002s.\*

6: Geneseq2003as.\*

7: Geneseq2003bs.\*

8: Geneseq2004s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	62	100.0	12	4	AAB61743 Human typ
2	62	100.0	25	7	AAB38632 Human C-t
3	62	100.0	262	7	AAB38640 Human col
4	62	100.0	270	2	AAY07375 Procollag
5	62	100.0	293	6	ABU70813 Human adi
6	62	100.0	309	4	AU23675 Novel hum
7	62	100.0	309	4	ABG60248 Human ova
8	62	100.0	309	5	ABG61719 Novel ova
9	62	100.0	339	6	ABU70735 Human adi
10	62	100.0	623	2	AAW12843 Pro-alpha
11	62	100.0	626	2	AAW12842 Truncated
12	62	100.0	842	7	ADE08475 Novel pro
13	62	100.0	1466	4	ABG50291 Collagen
14	62	100.0	1466	5	ABG90747 Human tum
15	62	100.0	1466	6	ABU54454 Human tum
16	62	100.0	1466	6	ABR47418 Breast ca
17	62	100.0	1469	4	ABG15191 Novel hum
18	62	100.0	1470	7	ADE09399 Novel pro
19	62	100.0	1726	6	ABR42661 Decorin-m
20	61	98.4	1466	4	AAB02537 Porcine a
21	56	90.3	11	4	AAB61742 Human typ
22	56	90.3	1078	2	AAR71704 Collagen
23	56	90.3	1078	3	AAY96125 Collagen
24	56	90.3	1078	5	AAB16478 Human col
25	56	90.3	1078	5	ABB80736 Collagen

## ALIGNMENTS

### RESULT 1

ID AAB61743 standard; peptide; 12 AA.

XX AAB61743;

DT 20-APR-2001 (first entry)

DE Human type III collagen carboxy-telopeptide fragment.

XX Type II collagen; immunoassay; cartilage; telopeptide; human;

KW type III collagen.

XX Synthetic.

OS Homo sapiens.

XX WO200079284-A1.

PD 28-DEC-2000.

PF 10-DEC-1999; 99WO-US029357.

PR 17-JUN-1999; 99US-00335098.

PR 29-JUN-1999; 99US-0141574P.

PR 02-JUL-1999; 99US-0142274P.

PR 07-JUL-1999; 99US-0142675P.

PR 30-AUG-1999; 99US-00385740.

PA (WASH-) WASHINGTON RES FOUND.

XX Eyre DR;

PI WPI; 2001-146859/15.

DR Assay for detecting cross-linked telopeptide analytes indicative of type

PT II collagen resorption in vivo in a body fluid sample, comprises

PT contacting the sample with an antibody which binds to the analyte.

XX Disclosure; Page 15; 34pp; English.

CC The invention relates to immunoassays for measuring type II collagen

CC (cartilage) resorption in vivo. The method of analysing a body fluid

CC sample for the presence of an analyte indicative of a physiological

CC condition, involves contacting the body fluid sample with an antibody

CC (Ab) which binds to the analyte, detecting binding of Ab in the body

CC fluid sample, and correlating any detected binding to the physiological

CC condition. The analysis is useful for measuring type II collagen

CC (cartilage) resorption in vivo, for distinguishing between resorption of

26 56 90.3 1078 5 ABB09628 Amino aci  
27 55 88.7 1466 4 AAE02534 Bovine al  
28 55 88.7 1466 4 AAE02533 Bovine al  
29 54 87.1 30 2 AAR92068 Recombina  
30 48 77.4 9 4 AAB61741 Human typ  
31 48 77.4 11 3 AAY68425 Type III  
32 48 77.4 11 3 AAY68436 Type III  
33 48 77.4 11 3 AAY82167 Type II a  
34 48 77.4 11 4 AAB51275 Collagen  
35 42 67.7 8 4 AAB61740 Human typ  
36 42 67.7 357 4 AEG08457 Novel hum  
37 41 66.1 176 4 AAU60152 Propionib  
38 41 66.1 176 6 ABM56671 Coryneb  
39 41 66.1 228 4 AAB76710 Coryneb  
40 41 66.1 691 6 ABU19644 Protein e  
41 40 64.5 133 4 AAO00933 Human pol  
42 40 64.5 383 2 AAY34469 Porphorym  
43 40 64.5 384 2 AAY34345 Porphorym  
44 40 64.5 438 2 AAW71488 Helicobac  
45 40 64.5 479 6 ABU23506 Protein e

CC non-mineralized and mineralized cartilage, and for measuring total  
 CC cartilage resorption in vivo. The present sequence represents a linear  
 CC synthetic peptide based on the carboxy-terminal telopeptide sequence of  
 CC human collagen type III  
 XX  
 SQ Sequence 12 AA;

Query Match 100.0%; Score 62; DB 4; Length 12;  
 Best Local Similarity 100.0%; Pred. No. 0.0028;  
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IAGIGGEKAGGF 12  
 |||||  
 Db 1 IAGIGGEKAGGF 12

RESULT 2  
 AAE38632  
 ID AAE38632 standard; peptide; 25 AA.  
 XX  
 AC AAE38632;  
 XX  
 DT 04-DEC-2003 (first entry)  
 XX  
 DE Human C-terminal telopeptide alpha 1 #4.  
 XX  
 KW Human; therapy; osteoporosis; left-ventricular hypertrophy; scleroderma;  
 KW liver cirrhosis; congestive heart failure; multiple sclerosis; fibrosis;  
 KW coal-miner's pneumoconiosis; muscle-wasting syndrome; diabetes mellitus;  
 KW rheumatoid arthritis; Alzheimer's disease; muscular dystrophy; cancer;  
 KW telopeptide alpha.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO2003068919-A2.  
 XX  
 PD 21-AUG-2003.  
 XX  
 PF 12-FEB-2003; 2003WO-US004183.  
 XX  
 PR 12-FEB-2002; 2002US-0356008P.  
 XX  
 PA (REGC ) UNIV CALIFORNIA.  
 XX  
 PI Hellerstein MK;  
 XX  
 WPI; 2003-689661/65.  
 XX  
 DR Determining rate of biosynthesis or breakdown of inaccessible biological  
 PT molecules, useful e.g. for diagnosis or monitoring treatment, by  
 PT administering labeled precursor.  
 XX  
 PS Claim 24; Page 99; 105pp; English.

CC The invention relates to a method of determining the rate of biosynthesis  
 CC or breakdown of at least one inaccessible biological molecule in a  
 CC subject. The method is useful for diagnosis or monitoring and treatment  
 CC of diseases associated with an altered rate of biosynthesis/breakdown of  
 CC an isotopically labelled precursor molecule, specifically osteoporosis;  
 CC left-ventricular hypertrophy; liver cirrhosis or fibrosis; congestive  
 CC heart failure; scleroderma; coal-miner's pneumoconiosis; cardiac or lung  
 CC fibrosis; Alzheimer's disease; multiple sclerosis; rheumatoid arthritis;  
 CC diabetes mellitus; muscle-wasting syndromes; muscular dystrophy; athletic  
 CC training and cancer. The method is also useful for screening candidate  
 CC gene or protein targets, phenotypic/human validation studies on potential  
 CC drugs, drug mechanism studies and determining the risk of developing the  
 CC disease. The present sequence is human C-terminal telopeptide alpha 1.  
 CC This sequence is used to illustrate the method of the invention  
 XX  
 SQ Sequence 25 AA;

Query Match 100.0%; Score 62; DB 7; Length 25;  
 Best Local Similarity 100.0%; Pred. No. 0.006;

Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 IAGIGGEKAGGF 12  
 |||||  
 Db 9 IAGIGGEKAGGF 20

RESULT 3  
 AAE38640  
 ID AAE38640 standard; protein; 262 AA.  
 XX  
 AC AAE38640;  
 XX  
 DT 04-DEC-2003 (first entry)  
 XX  
 DE Human collagen type I cross-linked C-terminal peptide, PIICP(alpha).  
 XX  
 KW Human; therapy; osteoporosis; left-ventricular hypertrophy; scleroderma;  
 KW liver cirrhosis; congestive heart failure; multiple sclerosis; fibrosis;  
 KW coal-miner's pneumoconiosis; muscle-wasting syndrome; diabetes mellitus;  
 KW rheumatoid arthritis; Alzheimer's disease; muscular dystrophy; cancer;  
 KW collagen type I cross-linked carboxy-terminal peptide.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO2003068919-A2.  
 XX  
 PD 21-AUG-2003.  
 XX  
 PF 12-FEB-2003; 2003WO-US004183.  
 XX  
 PR 12-FEB-2002; 2002US-0356008P.  
 XX  
 PA (REGC ) UNIV CALIFORNIA.  
 XX  
 PI Hellerstein MK;  
 XX  
 WPI; 2003-689661/65.

Determining rate of biosynthesis or breakdown of inaccessible biological  
 PT molecules, useful e.g. for diagnosis or monitoring treatment, by  
 PT administering labeled precursor.  
 XX  
 PS Claim 24; Page 102-103; 105pp; English.  
 XX  
 CC The invention relates to a method of determining the rate of biosynthesis  
 CC or breakdown of at least one inaccessible biological molecule in a  
 CC subject. The method is useful for diagnosis or monitoring and treatment  
 CC of diseases associated with an altered rate of biosynthesis/breakdown of  
 CC an isotopically labelled precursor molecule, specifically osteoporosis;  
 CC left-ventricular hypertrophy; liver cirrhosis or fibrosis; congestive  
 CC heart failure; scleroderma; coal-miner's pneumoconiosis; cardiac or lung  
 CC fibrosis; Alzheimer's disease; multiple sclerosis; rheumatoid arthritis;  
 CC diabetes mellitus; muscle-wasting syndromes; muscular dystrophy; athletic  
 CC training and cancer. The method is also useful for screening candidate  
 CC gene or protein targets, phenotypic/human validation studies on potential  
 CC drugs, drug mechanism studies and determining the risk of developing the  
 CC disease. The present sequence is human collagen type I cross-linked  
 CC carboxy terminal peptide (ICTP). This sequence is used to illustrate the  
 CC method of the invention  
 XX  
 SQ Sequence 262 AA;

Query Match 100.0%; Score 62; DB 7; Length 262;  
 Best Local Similarity 100.0%; Pred. No. 0.065;  
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IAGIGGEKAGGF 12  
 |||||  
 Db 1 IAGIGGEKAGGF 12

RESULT 4

AAU07375  
ID AAU07375 standard; protein; 270 AA.  
XX  
AC AAU07375;  
XX  
DT 16-JUL-1999 (first entry)  
XX  
DE Procollagen-III-C-terminal propeptide.  
XX  
XX Alcoholic cirrhosis; biliary cirrhosis; hepatitis; schistosomiasis;  
KW cardiac fibrosis; Crohn's disease; diabetic nephropathy; collagen;  
KW fibril; procollagen-III-C-terminal propeptide; immunoassay; diagnosis;  
KW glucocorticosteroid; antibody.  
XX  
XX Homo sapiens.  
OS  
PN EP913692-A1.  
XX  
PD 06-MAY-1999.  
XX  
PF 31-OCT-1997; 97EP-00119018.  
XX  
PR 31-OCT-1997; 97EP-00119018.  
XX  
PA (FARB ) BAYER AG.  
XX  
XX Burchardt ER, Kroll W, Neumann R, Schroeder W;  
PI WPI; 1999-256748/22.  
XX  
DR N-PSDB; AAX57501.  
XX

XX New immunoassay for procollagen-III-C-terminal propeptide (PIIICP) useful  
PT for diagnosing diseases such as alcoholic cirrhosis, hepatitis and  
PT Duchenne's muscular dystrophy.  
XX  
PS Disclosure; Fig 1; 28pp; English.  
XX  
XX Diseases such as alcoholic cirrhosis, biliary cirrhosis, hepatitis,  
CC schistosomiasis, cardiac fibrosis, Crohn's disease, diabetic nephropathy  
CC and fibroses caused by surgery result from an inappropriate production of  
CC collagen fibrils. Collagen III is synthesised as a preproprotein which is  
CC modified by post-translation modification. The mature collagen III  
CC fibrils are generated by cleavage of a propeptide from the C-terminus  
CC followed by cleavage of a fragment from the N-terminus. This sequence  
CC corresponds to the procollagen III C-terminal propeptide (PIIICP). As the  
CC fibrils are formed, the PIIICP propeptide is released into the  
CC surrounding tissue fluids. The invention relates to an immunoassay to  
CC determine PIIICP concentrations and thus diagnose the above diseases and  
CC others which result from inappropriate collagen deposition. Different  
CC fibrotic diseases may show different levels of PIIICP in the serum so the  
CC immunoassay may also be used to determine which disease a particular  
CC patient is suffering from. The new immunoassay may also be used to  
CC measure the rate of collagen synthesis in patients being treated with  
CC glucocorticosteroids. In addition the new antibodies may be used in  
CC immunohistochemical staining of cryostat and paraffin sections to assess  
CC collagen synthesis in tissue samples from patients suspected of having  
CC fibrotic disease  
XX  
SQ Sequence 270 AA;

Query Match 100.0%; Score 62; DB 2; Length 270;  
Best Local Similarity 100.0%; Pred. No. 0.067; Mismatches 0; Indels 0; Gaps 0;  
Matches 12; Conservative 0;

QY 1 IAGIGGKAGGF 12  
| | | | | | | | | |  
Db 9 IAGIGGKAGGF 20

RESULT 5  
ABU70813  
ID ABU70813 standard; protein; 293 AA.  
XX

ABU70813;  
XX  
DT 10-JUN-2003 (first entry)  
XX  
DE Human adipocyte Selected Interacting domain, SID, #444.  
XX  
KW Human; prey; adipocyte; SID; selected interacting domain; anorectic;  
KW antidiabetic; protein-protein interaction; diabetes;  
KW yeast 2-hybrid assay; metabolic disorder; obesity.  
XX  
OS Homo sapiens.  
XX  
PN WO200286122-A2.  
XX  
PD 31-OCT-2002.  
XX  
PF 14-MAR-2002; 2002WO-EP003768.  
XX  
PR 14-MAR-2001; 2001US-0275734P.  
XX  
PA (HYBR-) HYBRIGENICS.  
XX  
PI Legrain P, Daviet L;  
XX  
DR WPI; 2003-103412/09.  
XX  
DR N-PSDB; ACA57357.

XX New complex between two interacting proteins in adipocyte cells, useful  
PT for identifying selected interacting domains that modulate protein  
PT interactions, or for preventing or treating metabolic disorders such as  
PT obesity or diabetes.  
XX  
PS Claim 6; Page 258-259; 382pp; English.  
XX  
XX The invention relates to a complex between two interacting proteins in  
CC adipocyte cells, given in the specification. The proteins are identified  
CC by selecting a bait protein from a known adipocyte marker and then  
CC performing a yeast 2-hybrid selection to isolate prey proteins encoded by  
CC members of an adipocyte cDNA library. The proteins are designated SID  
CC (RTM) (selected interacting domains) proteins. Also included are a  
CC polynucleotide encoding a polypeptide in the adipocyte cells, a  
CC recombinant host cell expressing at least one of the interacting  
CC polypeptides of the complex, selecting a modulating compound in adipocyte  
CC cells, a SID (RTM) polypeptide comprising any of the 738 amino acid  
CC sequences given in the specification (including its fragment or variant),  
CC a SID (RTM) polynucleotide comprising any of the 738 nucleotide sequences  
CC given in the specification (including its fragment or variant), a vector  
CC comprising the SID (RTM) polynucleotide, a recombinant host cell  
CC comprising the vector, a protein chip comprising the polypeptides and a  
CC record comprising all or part of the data, listed in the specification.  
CC The complex, polypeptides, polynucleotides and compounds are useful for  
CC preventing or treating metabolic disorders such as obesity or diabetes.  
CC The polynucleotides are useful as probes or primers. The complex is  
CC particularly useful for identifying selected interacting domains (SID  
CC (RTM)) for screening drugs that modulate the protein interaction, thus  
CC exhibiting the therapeutic effect. The present sequence represents a SID  
CC (prey) protein of the invention  
XX  
SQ Sequence 293 AA;

Query Match 100.0%; Score 62; DB 6; Length 293;  
Best Local Similarity 100.0%; Pred. No. 0.073; Mismatches 0; Indels 0; Gaps 0;  
Matches 12; Conservative 0;

QY 1 IAGIGGKAGGF 12  
| | | | | | | | | |  
Db 136 IAGIGGKAGGF 147

RESULT 6  
AAU23675  
ID AAU23675 standard; protein; 309 AA.  
XX

AC AAU23675;  
XX  
DT 18-DEC-2001 (first entry)  
XX  
DE Novel human enzyme polypeptide #761.  
XX  
KW Human; oxidoreductase enzyme; transferase; hydrolase; lyase; isomerase;  
KW ligase; hyperproliferative disorder; immunodeficiency disorder;  
KW autoimmune disorder; neurological disorder; metabolic disorder;  
KW inflammatory disorder; cardiovascular disorder; reproductive disorder;  
KW blood-related disorder; infectious disorder; cytostatic; anti arthritic;  
KW nephrotropic; anticoagulant.  
XX  
OS Homo sapiens.  
XX  
FN WO20015301-A2.  
XX  
XX 02-AUG-2001.  
XX  
XX 17-JAN-2001; 2001WO-US001239.  
XX  
XX 31-JAN-2000; 2000US-0179065P.  
PR 04-FEB-2000; 2000US-0180628P.  
PR 24-FEB-2000; 2000US-0184664P.  
PR 02-MAR-2000; 2000US-0186350P.  
PR 16-MAR-2000; 2000US-0189874P.  
PR 17-MAR-2000; 2000US-0190076P.  
PR 18-APR-2000; 2000US-0198123P.  
PR 19-MAY-2000; 2000US-0205515P.  
PR 07-JUN-2000; 2000US-0209467P.  
PR 28-JUN-2000; 2000US-0214886P.  
PR 30-JUN-2000; 2000US-0215135P.  
PR 07-JUL-2000; 2000US-0216647P.  
PR 07-JUL-2000; 2000US-0216880P.  
PR 11-JUL-2000; 2000US-0217487P.  
PR 11-JUL-2000; 2000US-0217496P.  
PR 14-JUL-2000; 2000US-0218290P.  
PR 26-JUL-2000; 2000US-0220963P.  
PR 26-JUL-2000; 2000US-0220964P.  
PR 14-AUG-2000; 2000US-0224518P.  
PR 14-AUG-2000; 2000US-0224519P.  
PR 14-AUG-2000; 2000US-0225213P.  
PR 14-AUG-2000; 2000US-0225214P.  
PR 14-AUG-2000; 2000US-0225266P.  
PR 14-AUG-2000; 2000US-0225267P.  
PR 14-AUG-2000; 2000US-0225268P.  
PR 14-AUG-2000; 2000US-0225270P.  
PR 14-AUG-2000; 2000US-0225447P.  
PR 14-AUG-2000; 2000US-0225757P.  
PR 14-AUG-2000; 2000US-0225758P.  
PR 14-AUG-2000; 2000US-0225759P.  
PR 18-AUG-2000; 2000US-0226279P.  
PR 22-AUG-2000; 2000US-0226681P.  
PR 22-AUG-2000; 2000US-0226688P.  
PR 23-AUG-2000; 2000US-0227182P.  
PR 30-AUG-2000; 2000US-0227009P.  
PR 01-SEP-2000; 2000US-0228324P.  
PR 01-SEP-2000; 2000US-0228287P.  
PR 01-SEP-2000; 2000US-0229343P.  
PR 01-SEP-2000; 2000US-0229344P.  
PR 05-SEP-2000; 2000US-0229345P.  
PR 05-SEP-2000; 2000US-0229509P.  
PR 06-SEP-2000; 2000US-0229513P.  
PR 06-SEP-2000; 2000US-0230437P.  
PR 06-SEP-2000; 2000US-0230438P.  
PR 08-SEP-2000; 2000US-0231242P.  
PR 08-SEP-2000; 2000US-0231243P.  
PR 08-SEP-2000; 2000US-0231244P.  
PR 08-SEP-2000; 2000US-0231413P.  
PR 08-SEP-2000; 2000US-0231414P.  
PR 08-SEP-2000; 2000US-0232080P.  
PR 08-SEP-2000; 2000US-0232081P.  
PR 12-SEP-2000; 2000US-02331968P.  
PR  
PR 14-SEP-2000; 2000US-0232397P.  
PR 14-SEP-2000; 2000US-0232398P.  
PR 14-SEP-2000; 2000US-0232399P.  
PR 14-SEP-2000; 2000US-0232400P.  
PR 14-SEP-2000; 2000US-0232401P.  
PR 14-SEP-2000; 2000US-0233063P.  
PR 14-SEP-2000; 2000US-0233064P.  
PR 14-SEP-2000; 2000US-0233065P.  
PR 21-SEP-2000; 2000US-0234223P.  
PR 21-SEP-2000; 2000US-0234274P.  
PR 25-SEP-2000; 2000US-0234977P.  
PR 25-SEP-2000; 2000US-0234988P.  
PR 26-SEP-2000; 2000US-0235484P.  
PR 27-SEP-2000; 2000US-0235834P.  
PR 27-SEP-2000; 2000US-0235836P.  
PR 29-SEP-2000; 2000US-0236327P.  
PR 29-SEP-2000; 2000US-0236367P.  
PR 29-SEP-2000; 2000US-0236369P.  
PR 29-SEP-2000; 2000US-0236370P.  
PR 02-OCT-2000; 2000US-0236802P.  
PR 02-OCT-2000; 2000US-0237037P.  
PR 02-OCT-2000; 2000US-0237038P.  
PR 02-OCT-2000; 2000US-0237039P.  
PR 02-OCT-2000; 2000US-0237040P.  
PR 13-OCT-2000; 2000US-0239935P.  
PR 13-OCT-2000; 2000US-0239937P.  
PR 20-OCT-2000; 2000US-0240960P.  
PR 20-OCT-2000; 2000US-0241221P.  
PR 20-OCT-2000; 2000US-0241785P.  
PR 20-OCT-2000; 2000US-0241786P.  
PR 20-OCT-2000; 2000US-0241787P.  
PR 20-OCT-2000; 2000US-0241808P.  
PR 20-OCT-2000; 2000US-0241809P.  
PR 01-NOV-2000; 2000US-0244617P.  
PR 08-NOV-2000; 2000US-0246474P.  
PR 08-NOV-2000; 2000US-0246475P.  
PR 08-NOV-2000; 2000US-0246476P.  
PR 08-NOV-2000; 2000US-0246477P.  
PR 08-NOV-2000; 2000US-0246478P.  
PR 08-NOV-2000; 2000US-0246523P.  
PR 08-NOV-2000; 2000US-0246524P.  
PR 08-NOV-2000; 2000US-0246525P.  
PR 08-NOV-2000; 2000US-0246526P.  
PR 08-NOV-2000; 2000US-0246527P.  
PR 08-NOV-2000; 2000US-0246528P.  
PR 08-NOV-2000; 2000US-0246532P.  
PR 08-NOV-2000; 2000US-0246609P.  
PR 08-NOV-2000; 2000US-0246610P.  
PR 08-NOV-2000; 2000US-0246611P.  
PR 08-NOV-2000; 2000US-0246613P.  
PR 17-NOV-2000; 2000US-0249207P.  
PR 17-NOV-2000; 2000US-0249208P.  
PR 17-NOV-2000; 2000US-0249209P.  
PR 17-NOV-2000; 2000US-0249210P.  
PR 17-NOV-2000; 2000US-0249211P.  
PR 17-NOV-2000; 2000US-0249212P.  
PR 17-NOV-2000; 2000US-0249213P.  
PR 17-NOV-2000; 2000US-0249214P.  
PR 17-NOV-2000; 2000US-0249215P.  
PR 17-NOV-2000; 2000US-0249216P.  
PR 17-NOV-2000; 2000US-0249217P.  
PR 17-NOV-2000; 2000US-0249218P.  
PR 17-NOV-2000; 2000US-0249244P.  
PR 17-NOV-2000; 2000US-0249245P.  
PR 17-NOV-2000; 2000US-0249246P.  
PR 17-NOV-2000; 2000US-0249265P.  
PR 17-NOV-2000; 2000US-0249297P.  
PR 17-NOV-2000; 2000US-0249299P.  
PR 17-NOV-2000; 2000US-0249300P.  
PR 01-DEC-2000; 2000US-0250160P.  
PR 01-DEC-2000; 2000US-0250391P.

PR 05-DEC-2000; 2000US-0251030P.  
PR 05-DEC-2000; 2000US-0251988P.  
PR 06-DEC-2000; 2000US-0256719P.  
PR 06-DEC-2000; 2000US-0251479P.  
PR 08-DEC-2000; 2000US-0251856P.  
PR 08-DEC-2000; 2000US-0251868P.  
PR 08-DEC-2000; 2000US-0251869P.  
PR 08-DEC-2000; 2000US-0251989P.  
PR 08-DEC-2000; 2000US-0251990P.  
PR 11-DEC-2000; 2000US-0254097P.  
PR 05-JAN-2001; 2001US-0259678P.  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX  
XX Rosen CA, Barash SC, Ruben SM;  
XX  
XX WPI; 2001-465566/50.  
XX N-PSDB; AAS41545.  
XX  
XX Novel polypeptides and polynucleotides useful for diagnosing, preventing,  
PT treating neural, immune system, muscular, reproductive, pulmonary,  
PT cardiovascular, renal, proliferative disorders and cancerous diseases.  
XX  
XX Claim 11; SEQ ID NO 1671; 1180pp; English.  
XX  
XX The present invention relates to the isolation of novel human enzyme  
CC polypeptides, and the cDNA (AAS40785-AAS41684) and genomic sequences  
CC encoding them. The enzyme polypeptides of the invention may comprise the  
CC functional classes of oxidoreductases, transferases, hydrolases, lyases,  
CC isomerases or ligases. The sequences of the invention are useful in the  
CC diagnosis, treatment, prevention and/or prognosis of a wide range of  
CC disorders including hyperproliferative disorders (e.g. cancer),  
CC immunodeficiency disorders (e.g. AIDS) autoimmune disorders (e.g.  
CC arthritis), neurological disorders (e.g. Alzheimer's disease), metabolic  
CC disorders (e.g. phenylketonuria), inflammatory disorders (e.g. asthma),  
CC cardiovascular disorders (e.g. atherosclerosis), blood-related disorders  
CC (e.g. haemophilia), reproductive disorders (e.g. infertility) and  
CC infectious disorders (e.g. influenza). The polynucleotides of the  
CC invention can also be used in gene therapy. AAU22915-AAU23814 represent  
CC the novel human enzyme polypeptides of the invention. Note: The sequence  
CC data for this patent did not form part of the printed specification, but  
CC was obtained in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX  
XX SQ Sequence 309 AA;

Query Match 100.0%; Score 62; DB 4; Length 309;  
Best Local Similarity 100.0%; Pred. No. 0.077;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 IAGIGGEKAGGF 12  
Db 48 IAGIGGEKAGGF 59  
|||||  
|||||

RESULT 7  
ABG60248  
ID ABG60248 standard; protein; 309 AA.  
XX  
XX AC ABG60248;  
XX  
XX DT 13-AUG-2002 (first entry)  
XX  
XX DE Human ovarian antigen #10.  
XX  
XX KW Human; ovarian antigen; ovary disorder; breast disorder;  
KW neoplastic disorder; cancer; infectious disease; inflammatory disease;  
KW reproductive system disorder; autoimmune disorder; Alzheimer's disease;  
KW blood-related disorder; hyperproliferative disorder; hair loss;  
KW urinary system disorder; cardiovascular disorder; arrhythmia;  
KW respiratory disorder; musculoskeletal system disorder;  
KW neural activity disorder; neurological disorder; endocrine disorder;  
KW gastrointestinal disorder; liver disorder; pancreatic disorder;

KW gall bladder disorder; large intestine disorder; developmental disorder;  
KW inherited disorder; wound healing; skin aging; food additive;  
XX preservative.  
XX  
XX Homo sapiens.  
XX  
XX WO200155329-A2.  
XX  
XX 02-AUG-2001.  
XX  
XX 17-JAN-2001; 2001WO-US001360.  
XX  
XX 31-JAN-2000; 2000US-0179065P.  
XX 04-FEB-2000; 2000US-0180628P.  
XX 07-JUN-2000; 2000US-0209467P.  
XX 14-SEP-2000; 2000US-0232398P.  
XX 17-NOV-2000; 2000US-0249300P.  
XX 01-DEC-2000; 2000US-0250160P.  
XX 08-DEC-2000; 2000US-0251868P.  
XX 08-DEC-2000; 2000US-0251990P.  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX  
XX Rosen CA, Barash SC, Ruben SM;  
XX  
XX WPI; 2001-476195/51.  
XX N-PSDB; ABK72051.  
XX  
XX Novel isolated human ovarian related polypeptide useful for  
PT diagnosis/treatment of disorders of ovary and breast such as neoplastic  
PT disorders, infectious diseases, inflammatory diseases, and reproductive  
PT disorders.  
XX  
XX Claim 11; SEQ ID NO 78; 524pp; English.  
XX  
XX The invention relates to isolated ovarian related polypeptide (ovarian  
CC antigen) comprising a sequence at least 90% identical to a sequence  
CC selected from a polypeptide fragment, domain, epitope or full-length  
CC protein of a sequence (S1) appearing as ABG60239-ABG60296 having  
CC biological activity, or a variant, allelic variant or species homologue  
CC of S1. Also included are the cDNA clones encoding the proteins of S1. S1,  
CC an anti-S1 antibody and the cDNA are useful for diagnosing, preventing,  
CC treating or ameliorating a medical condition in mammalian subject  
CC especially diseases and/or disorders of the ovary and/or breast such as  
CC neoplastic diseases (such as ovarian Krukenberg tumour and cancer),  
CC infectious diseases (e.g., mastitis, oophoritis), inflammatory diseases  
CC (e.g., abscesses), reproductive system disorders (Paget's disease),  
CC autoimmune disorders (systemic lupus erythematosus, rheumatoid  
CC arthritis), blood-related disorders (sickle cell anaemia),  
CC hyperproliferative disorders, urinary system disorders  
CC (glomerulonephritis), cardiovascular disorders (arrhythmias), respiratory  
CC disorders, musculoskeletal system disorders, neural activity and  
CC neurological disorders (Alzheimer's disease and Parkinson's disease),  
CC endocrine disorders (Addison's disease), gastrointestinal disorders  
CC (inflammatory disorders), liver disorders (biliary liver cirrhosis),  
CC pancreatic and gall bladder disorders, disorders of the large intestine,  
CC developmental and inherited disorders, diseases at the cellular level,  
CC and wound healing and epithelial cell proliferation. They are also useful  
CC to prevent skin aging, for preventing hair loss, to maintain organs  
CC before transplantation or for supporting cell culture of primary tissues,  
CC to modulate mammalian characteristics such as body height, to modulate  
CC mammalian metabolism, to change a mammal's mental or physical state, and  
CC as food additive or preservative. The present sequence represents an  
CC ovarian antigen, S1 protein of the invention. Note: The sequence data for  
CC this patent did not form part of the printed specification, but was  
CC obtained in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences. Note: The sequence data for  
CC this patent did not form part of the printed specification, but was  
CC obtained in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX  
XX SQ Sequence 309 AA;

```
Query Match      100.0%; Score 62; DB 4; Length 309;
Best Local Similarity 100.0%; Pred.No. 0.077;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 IAGIGGKAGGF 12
Db      48 IAGIGGKAGGF 59

RESULT 8
ABG61719
ID ABG61719 standard; protein; 309 AA.
XX
AC
ABG61719;
XX
DT 26-AUG-2002 (first entry)
XX
DE
XX
XX Ovarian related polypeptide; neoplastic disorder; tumour; ovarian cancer;
KW hyperproliferative disorder; adult acute lymphocytic leukaemia;
KW breast cancer; reproductive system disorder; tuberculosis; arthritis;
KW immune system disorder; Chediak-Higashi's syndrome; neonatal neutropenia;
KW autoimmune disorder; Hashimoto's thyroiditis; inflammatory disorder;
KW septic shock; multiple sclerosis; central nervous system disorder;
KW neurological disorder; allergy; Parkinson's disease; Alzheimer's disease;
KW cardiovascular disorder; atherosclerosis; blood related disorder;
KW respiratory disorder; urinary system disorder; musculoskeletal disorder;
KW osteoporosis; wound healing; endocrine disorder; infectious disease;
KW gastrointestinal disorder; transplantation; food additive; preservative.
XX
OS Homo sapiens.
XX
XX US2002045230-A1.
XX
XX 18-APR-2002.
XX
XX 20-JUL-2001; 2001US-00908711.
XX
PR 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226868P.
PR 22-AUG-2000; 2000US-0227182P.
PR
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228224P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239353P.
PR 13-OCT-2000; 2000US-0239379P.
PR 20-OCT-2000; 2000US-0240360P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
```

PR 17-NOV-2000; 2000US-0249210P.  
PR 17-NOV-2000; 2000US-0249211P.  
PR 17-NOV-2000; 2000US-0249212P.  
PR 17-NOV-2000; 2000US-0249213P.  
PR 17-NOV-2000; 2000US-0249214P.  
PR 17-NOV-2000; 2000US-0249215P.  
PR 17-NOV-2000; 2000US-0249216P.  
PR 17-NOV-2000; 2000US-0249217P.  
PR 17-NOV-2000; 2000US-0249218P.  
PR 17-NOV-2000; 2000US-0249244P.  
PR 17-NOV-2000; 2000US-0249245P.  
PR 17-NOV-2000; 2000US-0249264P.  
PR 17-NOV-2000; 2000US-0249265P.  
PR 17-NOV-2000; 2000US-0249297P.  
PR 17-NOV-2000; 2000US-0249299P.  
PR 17-NOV-2000; 2000US-0249300P.  
PR 01-DEC-2000; 2000US-0250160P.  
PR 01-DEC-2000; 2000US-0250391P.  
PR 05-DEC-2000; 2000US-0251030P.  
PR 05-DEC-2000; 2000US-0251988P.  
PR 05-DEC-2000; 2000US-0256719P.  
PR 06-DEC-2000; 2000US-0251479P.  
PR 08-DEC-2000; 2000US-0251856P.  
PR 08-DEC-2000; 2000US-0251868P.  
PR 08-DEC-2000; 2000US-0251899P.  
PR 08-DEC-2000; 2000US-0251989P.  
PR 08-DEC-2000; 2000US-0251990P.  
PR 11-DEC-2000; 2000US-0254097P.  
PR 05-JAN-2001; 2001US-0259678P.  
PR 17-JAN-2001; 2001US-00764853.  
PR 17-JAN-2001; 2001US-00764856.  
PR 17-JAN-2001; 2001US-00764864.  
PR 17-JAN-2001; 2001US-00764867.  
PR 17-JAN-2001; 2001US-00764868.  
PR 17-JAN-2001; 2001US-00764869.  
PR 17-JAN-2001; 2001US-00764870.  
PR 17-JAN-2001; 2001US-00764874.  
PR 17-JAN-2001; 2001US-00764882.  
PR 17-JAN-2001; 2001US-00764888.  
PR 17-JAN-2001; 2001US-00764891.  
PR 17-JAN-2001; 2001US-00764892.  
PR 17-JAN-2001; 2001US-00764896.  
PR 17-JAN-2001; 2001US-00764898.  
PR 17-JAN-2001; 2001US-00764902.  
PR 17-JAN-2001; 2001US-00764905.  
PR 17-JAN-2001; 2001WO-US001239.  
PR 17-JAN-2001; 2001WO-US001307.  
PR 17-JAN-2001; 2001WO-US001312.  
PR 17-JAN-2001; 2001WO-US001320.  
PR 17-JAN-2001; 2001WO-US001329.  
PR 17-JAN-2001; 2001WO-US001334.  
PR 17-JAN-2001; 2001WO-US001336.  
PR 17-JAN-2001; 2001WO-US001339.  
PR 17-JAN-2001; 2001WO-US001340.  
PR 17-JAN-2001; 2001WO-US001341.  
PR 17-JAN-2001; 2001WO-US001344.  
PR 17-JAN-2001; 2001WO-US001345.  
PR 17-JAN-2001; 2001WO-US001347.  
PR 17-JAN-2001; 2001WO-US001348.  
PR 17-JAN-2001; 2001WO-US001360.  
XX (ROSE/) ROSEN C A.  
PA (RUBI/) RUBIN S M.  
PA (BARA/) BARASH S C.  
XX

Query Match 100.0%; Score 62; DB 5; Length 309;  
Best Local Similarity 100.0%; Pred. No. 0.077;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 IAGIGGKAGGF 12  
DB 48 IAGIGGKAGGF 59

RESULT 9  
ABU70735  
ID ABU70735 standard; protein; 339 AA.  
XX  
AC ABU70735;  
XX  
DT 10-JUN-2003 (first entry)  
XX  
DE Human adipocyte Selected Interacting domain, SID, #366.  
XX  
KW Human; prey; adipocyte; SID; selected interacting domain; anorectic;  
KW anti-diabetic; protein-protein interaction; diabetes;  
KW yeast 2-hybrid assay; metabolic disorder; obesity.  
XX  
OS Homo sapiens.  
XX  
PN W0200286122-A2.  
XX  
PD 31-OCT-2002.  
XX  
PF 14-MAR-2002; 2002WO-EP003768.  
XX  
PR 14-MAR-2001; 2001US-0275734P.  
XX (HYBR-) HYBRIGENICS.  
XX  
PI Legrain P, Daviet L;  
XX  
DR WPI; 2003-103412/09.  
DR N-PSDB; ACAS7279.  
XX  
XX New complex between two interacting proteins in adipocyte cells, useful  
PT for identifying selected interacting domains that modulate protein  
PT interactions, or for preventing or treating metabolic disorders such as  
PT obesity or diabetes.  
XX  
PS Claim 6; Page 229; 382pp; English.  
XX  
CC The invention relates to a complex between two interacting proteins in  
CC adipocyte cells, given in the specification. The proteins are identified  
CC by selecting a bait protein from a known adipocyte marker and then  
CC performing a yeast 2-hybrid selection to isolate prey proteins encoded by  
CC members of an adipocyte cDNA library. The proteins are designated SID  
CC (RTM) (selected interacting domains) proteins. Also included are a  
CC polynucleotide encoding a polypeptide in the adipocyte cells, a  
CC recombinant host cell expressing at least one of the interacting  
CC polypeptides of the complex, selecting a modulating compound in adipocyte  
CC cells, a SID (RTM) polypeptide comprising any of the 738 amino acid  
CC sequences given in the specification (including its fragment or variant),  
CC a SID (RTM) polynucleotide comprising any of the 738 nucleotide sequences  
CC given in the specification (including its fragment or variant), a vector  
CC comprising the SID (RTM) polynucleotide, a recombinant host cell  
CC comprising the vector, a protein chip comprising the polypeptides and a  
CC record comprising all or part of the data, listed in the specification.  
CC The complex, polypeptides, polynucleotides and compounds are useful for  
CC preventing or treating metabolic disorders such as obesity or diabetes.  
CC The polynucleotides are useful as probes or primers. The complex is  
CC particularly useful for identifying selected interacting domains (SID  
CC (RTM)) for screening drugs that modulate the protein interaction, thus  
CC exhibiting the therapeutic effect. The present sequence represents a SID  
CC (prey) protein of the invention  
XX  
SQ Sequence 339 AA;

Query Match 100.0%; Score 62; DB 6; Length 339;  
Best Local Similarity 100.0%; Pred. No. 0.084;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 IAGIGGKAGGF 12  
DB 78 IAGIGGKAGGF 89

```

RESULT 10
AAW12843
ID AAW12843 standard; peptide; 623 AA.
XX
XX AAW12843;
AC
XX
XX
DT 15-DEC-1997 (first entry)
XX
DE Pro-alpha(III):(I) CP chimeric protein.
XX
XX C-propeptide; recognition sequence; procollagen; monomer chain; therapy;
KW trimerised pro-alpha chain; fibril; procollagen suicide; wound healing;
KW fibrotic disease; human; chimeric protein.
XX
XX Homo sapiens.
OS
XX
XX WO9708311-A1.
PN
XX
XX 06-MAR-1997.
PD
XX
XX 30-AUG-1996; 96WO-GB002122.
PF
XX
XX 31-AUG-1995; 95GB-00017773.
PR
XX 23-MAR-1996; 96GB-00006152.
PR
XX 14-JUN-1996; 96GB-00012476.
PR
XX
XX (UYMA-) UNIV VICTORIA MANCHESTER.
PA
XX
XX Bulleid N, Kadler K;
PI
XX WPI; 1997-179268/16.
DR
XX
XX Novel pro:collagen mol. - comprising pro:collagen C-pro:peptide attached
PT to an alien collagen alpha-chain or non-collagen material, useful e.g.
PT for wound healing.
XX
XX Example 1; Page 32-35; 69pp; English.
XX
XX This sequence represents a chimeric procollagen molecule of the
CC invention. This sequence has the procollagen C-propeptide from the pro
CC -alpha2(I) chain attached to the pro-alpha(III) chain sequence. The C-
CC propeptide is implicated in the assembly of the monomer chains into
CC trimerised pro-alpha chains prior to cleavage of the N- and C-propeptides
CC and formation of collagen in fibril-forming pro-alpha chains. The C-
CC propeptides determine the type-specific assembly of the moieties to which
CC they are attached. The molecule of the invention comprises a first moiety
CC having procollagen C-propeptide activity attached to a second moiety,
CC which is an alien collagen alpha-chain or a non-collagen material. The
CC novel collagen molecule can be used for treatment or diagnosis in humans
CC or animals, especially for the treatment of procollagen suicide, as an
CC adhesive or implant, to promote (chronic) wound healing or fibrotic
CC diseases with reduced scarring or for use in photography, brewing,
CC foodstuffs or textiles. The novel collagen molecules, especially when
CC containing substitutions in the recognition site, may have significantly
CC altered properties and characteristics, such as different binding
CC kinetics or alpha-chain selection properties
XX
XX Sequence 623 AA;
SQ
Query Match 100.0%; Score 62; DB 2; Length 623;
Best Local Similarity 100.0%; Pred. No. 0.16;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 IAGIGGEKAGGF 12
|||
Db 365 IAGIGGEKAGGF 376

RESULT 11
AAW12842
ID AAW12842 standard; protein; 626 AA.

```

```

XX
AC AAW12842;
XX
DT 15-DEC-1997 (first entry)
XX
XX Truncated pro-alpha(III) chain.
DE
XX
XX C-propeptide; recognition sequence; procollagen; monomer chain; therapy;
KW trimerised pro-alpha chain; fibril; procollagen suicide; wound healing;
KW fibrotic disease; human.
XX
XX Homo sapiens.
OS
XX
XX WO9708311-A1.
PN
XX
XX 06-MAR-1997.
PD
XX
XX 30-AUG-1996; 96WO-GB002122.
PF
XX
XX 31-AUG-1995; 95GB-00017773.
PR
XX 23-MAR-1996; 96GB-00006152.
PR
XX 14-JUN-1996; 96GB-00012476.
PR
XX
XX (UYMA-) UNIV VICTORIA MANCHESTER.
PA
XX
XX Bulleid N, Kadler K;
PI
XX WPI; 1997-179268/16.
DR
XX N-PSDB; AAT59892.
XX
XX Novel pro:collagen mol. - comprising pro:collagen C-pro:peptide attached
PT to an alien collagen alpha-chain or non-collagen material, useful e.g.
PT for wound healing.
XX
XX Example 1; Page 28-31; 69pp; English.
XX
XX This sequence represents a truncated procollagen pro-alpha(III) chain
CC that can be used in the procollagen molecules of the invention. The C-
CC propeptide is implicated in the assembly of the monomer chains into
CC trimerised pro-alpha chains prior to cleavage of the N- and C-propeptides
CC and formation of collagen in fibril-forming pro-alpha chains. The C-
CC propeptides determine the type-specific assembly of the moieties to which
CC they are attached. The molecule of the invention comprises a first moiety
CC having procollagen C-propeptide activity attached to a second moiety,
CC which is an alien collagen alpha-chain or a non-collagen material. The
CC novel collagen molecule can be used for treatment or diagnosis in humans
CC or animals, especially for the treatment of procollagen suicide, as an
CC adhesive or implant, to promote (chronic) wound healing or fibrotic
CC diseases with reduced scarring or for use in photography, brewing,
CC foodstuffs or textiles. The novel collagen molecules, especially when
CC containing substitutions in the recognition site, may have significantly
CC altered properties and characteristics, such as different binding
CC kinetics or alpha-chain selection properties
XX
XX Sequence 626 AA;
SQ
Query Match 100.0%; Score 62; DB 2; Length 626;
Best Local Similarity 100.0%; Pred. No. 0.16;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 IAGIGGEKAGGF 12
|||
Db 365 IAGIGGEKAGGF 376

RESULT 12
ADE08475
ID ADE08475 standard; protein; 842 AA.
XX
XX ADE08475;
AC
XX
XX 29-JAN-2004 (first entry)
DT
XX

```



DE Novel protein (useful for identifying genetic disorders) #630.

XX novel gene; novel protein; tissue marker; molecular weight marker;  
KW chromosome marker; genetic disorder.

XX Unidentified.

XX WO2003054152-A2.

XX 03-JUL-2003.

XX 10-DEC-2002; 2002WO-US039555.

XX 10-DEC-2001; 2001US-0339739P.

XX 11-DEC-2001; 2001US-0339453P.

XX 14-MAR-2002; 2002US-0365091P.

XX 14-MAR-2002; 2002US-0365384P.

XX 12-APR-2002; 2002US-0372381P.

XX 12-APR-2002; 2002US-0372615P.

XX 22-APR-2002; 2002US-00128558.

XX 24-APR-2002; 2002US-0376045P.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Asundi V, Goodrich RW, Ren F, Zhang J, Zhao QA, Wang J;  
PI Ghosh M, Xue AJ, Wehrman T, Weng G, Zhou P, Drmanac RT, Wang Z;  
PI Ma Y, Wang D, Chen R, Xu C, Boyle BJ;

DR WPI; 2003-569235/53.

DR N-PSDB; ADE07564.

XX New polynucleotides, useful for expressing recombinant proteins for  
PT analysis, characterization or therapeutic use, or as markers for tissues  
PT in which the corresponding protein is preferentially expressed.

XX Claim 20; SEQ ID NO 1541; 1177pp; English.

XX The invention comprises the amino acid and coding sequences of novel  
CC proteins. The DNA and protein sequences of the invention are useful as:  
CC markers for tissues in which the corresponding protein is preferentially  
CC expressed; as molecular weight markers on gels; as chromosome markers or  
CC tags; to identify chromosomes or to map related gene positions; and to  
CC compare with endogenous DNA sequences in patients to identify potential  
CC genetic disorders. The present amino acid sequence represents a protein  
CC of the invention.

XX Sequence 842 AA;

Query Match 100.0%; Score 62; DB 7; Length 842;  
Best Local Similarity 100.0%; Pred. No. 0.21; Mismatches 0; Indels 0; Gaps 0;  
Matches 12; Conservative 0;

Oy 1 IAGIGGEKAGGF 12  
|||||

Db 581 IAGIGGEKAGGF 592  
|||||

RESULT 13

ABB50291

ID ABB50291 standard; protein; 1466 AA.

XX ABB50291;

AC ABB50291;

XX 08-FEB-2002 (first entry)

XX Collagen type III alpha-1 ovarian tumour marker protein, SEQ ID NO:72.

XX Ovarian tumour marker gene; human; overexpression; upregulation;  
KW epithelial tumour; cancer; diagnosis; prognosis; disease monitoring;  
KW identification; serous cystadenoma; borderline serous tumour;  
KW serous cystadenocarcinoma; mucinous cystadenocarcinoma;  
KW mucinous cystadenoma; borderline mucinous tumour; endometrioid carcinoma;  
KW undifferentiated carcinoma; clear cell adenocarcinoma; cystadenofibroma;

adenofibroma; Brenner tumour; serial analysis of gene expression; SAGE;  
immune response pathway; cell proliferation regulation; protein folding;  
membrane localised; secreted; therapeutic target; cytostatic;  
gene therapy; vaccine.

XX Homo sapiens.

XX WO200175177-A2.

XX 11-OCT-2001.

XX 03-APR-2001; 2001WO-US010947.

XX 03-APR-2000; 2000US-0194336P.

XX (USSS) US DEPT HEALTH & HUMAN SERVICES.

XX Morin PJ, Sherman-Baust CA, Pizer ES, Hough CD;

XX WPI; 2001-6264450/72.

XX N-PSDB; ABA83117.

XX Detecting and identifying ovarian tumor, identifying increased risk for  
PT developing ovarian cancer, and determining effectiveness of ovarian  
PT cancer treatment, by measuring expression level of ovarian tumor marker  
PT gene.

XX Claim 23; Page 114-117; 140pp; English.

XX The invention relates to methods for diagnosing and prognosing ovarian  
CC tumours in an individual via the detection and measurement of the  
CC expression of ovarian tumour marker genes (ABA83081-ABA83122, ABA83180,  
CC ABA83182 and ABA83184) or segments thereof (ABA83123-ABA83189, ABA83179,  
CC ABA83181 and ABA83183). The methods of the invention are useful for  
CC detecting an ovarian tumour in a patient, for identifying an individual  
CC at increased risk for developing ovarian cancer, in prognostic tests for  
CC assessing the relative severity of ovarian cancer, in tests for  
CC monitoring a patient in remission from ovarian cancer and in tests for  
CC monitoring disease status in a patient being treated for ovarian cancer.  
CC The methods can additionally be used to identify a particular tumour as  
CC being an ovarian tumour (i.e., an epithelial ovarian tumour selected from  
CC serous cystadenoma, borderline serous tumour, serous cystadenocarcinoma,  
CC mucinous cystadenoma, borderline mucinous tumour, mucinous  
CC cystadenocarcinoma, endometrioid carcinoma, undifferentiated carcinoma,  
CC clear cell adenocarcinoma, cystadenofibroma, adenofibroma and Brenner  
CC tumour. The ovarian tumour marker genes of the invention were identified  
CC using SAGE (serial analysis of gene expression) and were found to be  
CC overexpressed in a broad variety of ovarian epithelial tumour cells  
CC relative to normal ovarian epithelial cells. The marker genes are  
CC implicated in immune response pathways, in the regulation of cell  
CC proliferation and in protein folding, and many of these are membrane-  
CC localised or secreted. In addition to their use as diagnostic and  
CC prognostic markers, the ovarian tumour marker genes or their encoded  
CC proteins may be used as therapeutic targets for the treatment and  
CC prevention of ovarian cancer. Sequences ABB50257-ABB50299 represent  
CC proteins encoded by ovarian tumour marker genes of the invention

XX Sequence 1466 AA;

Query Match 100.0%; Score 62; DB 4; Length 1466;  
Best Local Similarity 100.0%; Pred. No. 0.37; Mismatches 0; Indels 0; Gaps 0;  
Matches 12; Conservative 0;

Oy 1 IAGIGGEKAGGF 12  
|||||

Db 1205 IAGIGGEKAGGF 1216  
|||||

RESULT 14

ABB90747

ID ABB90747 standard; protein; 1466 AA.

XX ABB90747;

```
XX 30-MAY-2002 (first entry)
DE Human Tumour Endothelial Marker polypeptide SEQ ID NO 226.
XX Human; mouse; rat; TEM; tumour endothelial marker; NEM; PEM; cytostatic;
XX normal endothelial marker; pan-endothelial marker; immunostimulant;
KW antiangiogenic; tumour; neoangiogenesis; vascularised tumour;
KW polycystic kidney disease; diabetes; retinopathy; rheumatoid arthritis;
KW psoriasis.
XX Homo sapiens.
OS
XX WO200210217-A2.
PN
XX 07-FEB-2002.
PD
XX 01-AUG-2001; 2001WO-US024031.
PF
XX 02-AUG-2000; 2000US-0222599P.
PR
XX 11-AUG-2000; 2000US-0224360P.
PR
XX 11-APR-2001; 2001US-0282850P.
PR
XX (UYJO ) UNIV JOHNS HOPKINS.
PA
XX St Croix B, Kinzler KW, Vogelstein B;
PI
XX WPI; 2002-291856/33.
DR
XX N-PSDB; ABL92101.
XX
XX An isolated molecule comprising an antibody variable region which
PT specifically binds to an extracellular domain of a tumor endothelial
PT marker (TEM) protein, useful for inhibiting tumor growth.
XX
XX Claim 68; Page 197-200; 331pp; English.
XX
XX The invention relates to an isolated molecule comprising an antibody
CC variable region which specifically binds to an extracellular domain of a
CC tumor endothelial marker (TEM) protein selected from ABB90732, ABB90740,
CC ABB90749, ABB90750 and ABB90769. The antibodies which bind to TEM
CC proteins have cytostatic, immunostimulant and antiangiogenic activity.
CC They are useful for inhibiting tumor growth, neoangiogenesis in subjects
CC bearing a vascularised tumour, polycystic kidney disease, diabetic
CC retinopathy, rheumatoid arthritis and psoriasis. Human, mouse and rat TEM
CC genes and the encoded proteins (ABL92075-ABL92141 and ABB90721-ABB90789)
CC are disclosed, as are marker oligonucleotide sequences: tumour
CC endothelial markers (TEM) ABL91996-ABU92041 and ABL92143-ABL92191; normal
CC endothelial markers (NEM) ABL92042-ABU92074; and pan-endothelial markers
CC (PEM) ABL91903-ABL91995
XX
XX Sequence 1466 AA;
SQ
Query Match 100.0%; Score 62; DB 5; Length 1466;
Best Local Similarity 100.0%; Pred. No. 0.37;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 IAGIGGEKAGGF 12
DB 1205 IAGIGGEKAGGF 1216
RESULT 15
ID ABU54454
XX ABU54454 standard; protein; 1466 AA.
XX
XX ABU54454;
AC
XX 12-MAR-2003 (first entry)
DT
XX Human tumour endothelial marker TEM 15.
DE
XX Human; endothelial cell; EC; tumour endothelial cell; TEM; NEM;
KW Tumour endothelial marker; normal endothelial marker; PEM;
```

---

```
KW pan-endothelial marker; polycystic kidney disease; psoriasis;
KW diabetic retinopathy; rheumatoid arthritis; tumour angiogenesis;
KW neoangiogenesis; immune response; cytostatic; anti-diabetic;
KW ophthalmological; anti-rheumatic; anti-arthritis; anti-psoriatic.
XX
XX Homo sapiens.
OS
XX WO200283874-A2.
PN
XX 24-OCT-2002.
PD
XX 10-APR-2002; 2002WO-US008253.
PF
XX 11-APR-2001; 2001US-0282850P.
PR
XX 06-FEB-2002; 2002US-0354262P.
PR
XX (UYJO ) UNIV JOHNS HOPKINS.
PA
XX Carson-Walter E, St Croix B, Kinzler KW, Vogelstein B;
PI
XX WPI; 2003-093016/08.
DR
XX N-PSDB; ABX72026.
XX
XX New purified human transmembrane protein, designated as tumor endothelial
PT marker (TEM) 3, useful for detecting, diagnosing or treating tumors,
PT polycystic kidney disease, diabetic retinopathy, rheumatoid arthritis or
PT psoriasis.
XX
XX Disclosure; Page 211-214; 374pp; English.
XX
XX The present invention relates to a novel method for the isolation of
CC endothelial cells (ECs), and the identification of genes expressed in
CC normal and tumour ECs. Tumour endothelial marker (TEM), normal
CC endothelial marker (NEM), and pan-endothelial marker (PEM) genes are
CC identified in human ECs. The human EC marker proteins and the
CC polynucleotide sequences encoding them are useful for detecting,
CC diagnosing or treating tumors as well as polycystic kidney disease,
CC diabetic retinopathy, rheumatoid arthritis, and psoriasis. They are also
CC useful for inhibiting neoangiogenesis or tumour angiogenesis, for
CC inducing an immune response to tumour endothelial cells in a patient, or
CC for identifying candidate drugs for treating tumours. The present
CC sequence represents a human TEM or NEM protein of the invention
XX
XX Sequence 1466 AA;
SQ
Query Match 100.0%; Score 62; DB 6; Length 1466;
Best Local Similarity 100.0%; Pred. No. 0.37;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 IAGIGGEKAGGF 12
DB 1205 IAGIGGEKAGGF 1216
Search completed: September 18, 2004, 04:25:33
Job time : 65.2353 secs
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 18, 2004, 04:20:54 ; Search time 19.7647 Seconds  
(without alignments)  
31.344 Million cell updates/sec

Title: US-10-615-959-43  
Perfect score: 62  
Sequence: 1 IAGIGGEKAGGF 12

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA:\*

1: /cgn2\_6/ptodata/2/iaa/5A.COMB.pdp:\*  
2: /cgn2\_6/ptodata/2/iaa/5B.COMB.pdp:\*  
3: /cgn2\_6/ptodata/2/iaa/6A.COMB.pdp:\*  
4: /cgn2\_6/ptodata/2/iaa/6B.COMB.pdp:\*  
5: /cgn2\_6/ptodata/2/iaa/PCUTS.COMB.pdp:\*  
6: /cgn2\_6/ptodata/2/iaa/backfiles1.pdp:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	62	100.0	12	4	US-10-009-999A-43
2	62	100.0	623	3	Sequence 43, Appli
3	62	100.0	626	3	Sequence 3, Appli
4	56	90.3	11	4	US-09-029-348-2
5	56	90.3	1078	3	Sequence 2, Appli
6	56	90.3	1078	4	Sequence 42, Appl
7	56	90.3	1078	4	Sequence 21, Appl
8	56	90.3	1078	4	Sequence 21, Appl
9	54	87.1	15	4	Sequence 21, Appl
10	48	77.4	9	4	Sequence 20, Appl
11	48	77.4	11	6	Sequence 41, Appl
12	44	71.0	227	4	Patent No. 5473052
13	42	67.7	8	4	Sequence 30541, A
14	41	66.1	221	4	Sequence 40, Appl
15	41	66.1	503	4	Sequence 28805, A
16	40	64.5	674	4	Sequence 28805, A
17	39	62.9	157	2	Sequence 23643, A
18	39	62.9	247	4	Sequence 2476, A
19	39	62.9	247	4	Sequence 4, Appli
20	39	62.9	591	4	Sequence 363, App
21	38	62.9	674	4	Sequence 21412, A
22	38	61.3	319	3	Sequence 12837, A
23	38	61.3	319	3	Sequence 51, Appl
24	38	61.3	401	6	Sequence 12, Appl
25	38	61.3	419	4	Patent No. 5252556-1
26	38	61.3	490	4	Sequence 7295, Ap
27	38	61.3	576	4	Sequence 22715, A
					Sequence 32499, A

28	38	61.3	725	4	US-09-252-991A-27825	Sequence 27825, A
29	38	61.3	968	4	US-09-252-991A-18925	Sequence 18925, A
30	37	59.7	219	1	US-08-164-614A-9	Sequence 9, Appli
31	37	59.7	219	2	US-08-456-489B-9	Sequence 9, Appli
32	37	59.7	227	4	US-09-252-991A-30027	Sequence 30027, A
33	37	59.7	310	4	US-09-252-991A-19986	Sequence 19986, A
34	37	59.7	336	4	US-09-252-991A-21790	Sequence 21790, A
35	37	59.7	379	1	US-08-164-614A-8	Sequence 8, Appli
36	37	59.7	379	2	US-08-456-489B-8	Sequence 8, Appli
37	37	59.7	463	4	US-09-252-991A-20486	Sequence 20486, A
38	37	59.7	468	1	US-08-164-614A-7	Sequence 7, Appli
39	37	59.7	468	2	US-08-456-489B-7	Sequence 7, Appli
40	37	59.7	555	4	US-09-252-991A-27809	Sequence 27809, A
41	37	59.7	811	4	US-09-252-991A-28570	Sequence 28570, A
42	37	59.7	927	4	US-09-252-991A-16765	Sequence 16765, A
43	37	59.7	1084	4	US-09-252-991A-27188	Sequence 27188, A
44	37	59.7	1504	4	US-09-252-991A-26608	Sequence 26608, A
45	36.5	58.9	851	4	US-09-252-991A-22021	Sequence 22021, A

ALIGNMENTS

RESULT 1

US-10-009-999A-43

; Sequence 43, Application US/10009999A

; Patent No. 6602980

; GENERAL INFORMATION:

; APPLICANT: Eyre, David R.

; TITLE OF INVENTION: SYNTHETIC PEPTIDES FOR COLLAGEN RESORPTION ASSAYS

; FILE REFERENCE: WPOS-1-18220

; CURRENT APPLICATION NUMBER: US/10/009,999A

; CURRENT FILING DATE: 2003-01-22

; PRIOR APPLICATION NUMBER: PCT/US99/29357

; PRIOR FILING DATE: 1999-12-10

; PRIOR APPLICATION NUMBER: US 09/335,098

; PRIOR FILING DATE: 1999-06-17

; PRIOR APPLICATION NUMBER: US 60/141,574

; PRIOR FILING DATE: 1999-06-29

; PRIOR APPLICATION NUMBER: US 60/142,274

; PRIOR FILING DATE: 1999-07-02

; PRIOR APPLICATION NUMBER: US 60/142,675

; PRIOR FILING DATE: 1999-07-07

; PRIOR APPLICATION NUMBER: US 09/385,740

; PRIOR FILING DATE: 1999-08-30

; NUMBER OF SEQ ID NOS: 45

; SOFTWARE: PatentIn version 3.2

; SEQ ID NO 43

; LENGTH: 12

; TYPE: PRT

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: MISC FEATURE

; OTHER INFORMATION: synthetic peptide corresponding to carboxy-terminal

; OTHER INFORMATION: telopeptide sequence of human type III collagen

US-10-009-999A-43

Query Match 100.0%; Score 62; DB 4; Length 12;  
Best Local Similarity 100.0%; Pred. No. 0.00088;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 IAGIGGEKAGGF 12  
|  
Db 1 IAGIGGEKAGGF 12

RESULT 2

US-09-029-348-3

; Sequence 3, Application US/09029348

; Patent No. 6171827

; GENERAL INFORMATION:

; APPLICANT: THE VICTORIA UNIVERSITY OF MANCHESTER

; TITLE OF INVENTION: NOVEL PROCOLLAGENS

; FILE REFERENCE: d087857PUS LISTING  
; CURRENT APPLICATION NUMBER: US/09/029,348  
; CURRENT FILING DATE: 1998-05-07  
; NUMBER OF SEQ ID NOS: 20  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 3  
; LENGTH: 623  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: SEQUENCE  
; OTHER INFORMATION: DERIVED FROM CDNA OF PROCOLLAGENS  
US-09-029-348-3

Query Match 100.0%; Score 62; DB 3; Length 623;  
Best Local Similarity 100.0%; Pred. No. 0.044;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IAGIGGEKAGGF 12  
|||  
Db 365 IAGIGGEKAGGF 376

## RESULT 3

US-09-029-348-2  
; Sequence 2, Application US/09029348  
; Patent No. 6171827  
; GENERAL INFORMATION:  
; APPLICANT: THE VICTORIA UNIVERSITY OF MANCHESTER  
; TITLE OF INVENTION: NOVEL PROCOLLAGENS  
; FILE REFERENCE: d087857PUS LISTING  
; CURRENT APPLICATION NUMBER: US/09/029,348  
; CURRENT FILING DATE: 1998-05-07  
; NUMBER OF SEQ ID NOS: 20  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 2  
; LENGTH: 626  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: SEQUENCE  
; OTHER INFORMATION: DERIVED FROM CDNA OF PROCOLLAGENS  
US-09-029-348-2

Query Match 100.0%; Score 62; DB 3; Length 626;  
Best Local Similarity 100.0%; Pred. No. 0.044;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IAGIGGEKAGGF 12  
|||  
Db 365 IAGIGGEKAGGF 376

## RESULT 4

US-10-009-999A-42  
; Sequence 42, Application US/10009999A  
; Patent No. 6602980  
; GENERAL INFORMATION:  
; APPLICANT: Eyre, David R.  
; TITLE OF INVENTION: SYNTHETIC PEPTIDES FOR COLLAGEN RESORPTION ASSAYS  
; FILE REFERENCE: WROS-1-18220  
; CURRENT APPLICATION NUMBER: US/10/009,999A  
; CURRENT FILING DATE: 2003-01-22  
; PRIOR APPLICATION NUMBER: PCT/US99/29357  
; PRIOR FILING DATE: 1999-12-10  
; PRIOR APPLICATION NUMBER: US 09/335,098  
; PRIOR FILING DATE: 1999-06-17  
; PRIOR APPLICATION NUMBER: US 60/141,574  
; PRIOR FILING DATE: 1999-06-29  
; PRIOR APPLICATION NUMBER: US 60/142,274  
; PRIOR FILING DATE: 1999-07-02  
; PRIOR APPLICATION NUMBER: US 60/142,675  
; PRIOR FILING DATE: 1999-07-07

; PRIOR APPLICATION NUMBER: US 09/385,740  
; PRIOR FILING DATE: 1999-08-30  
; NUMBER OF SEQ ID NOS: 45  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 42  
; LENGTH: 11  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: MISC\_FEATURE  
; OTHER INFORMATION: synthetic peptide corresponding to carboxy-terminal  
; OTHER INFORMATION: telopeptide sequence of human type III collagen  
US-10-009-999A-42

Query Match 90.3%; Score 56; DB 4; Length 11;  
Best Local Similarity 100.0%; Pred. No. 0.007;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IAGIGGEKAGG 11  
|||  
Db 1 IAGIGGEKAGG 11

## RESULT 5

US-08-963-825-21  
; Sequence 21, Application US/08963825  
; Patent No. 6110689  
; GENERAL INFORMATION:  
; APPLICANT: Qvist, Per  
; TITLE OF INVENTION: A Method for Assaying Collagen Fragments  
; TITLE OF INVENTION: in Body Fluids, A Test Kit and Means for Carrying Out the  
; TITLE OF INVENTION: Method and Use of the Method to Diagnose the Presence of  
; TITLE OF INVENTION: Disorders Associated with the Metabolism of  
; NUMBER OF SEQUENCES: 21  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Darby & Darby PC  
; STREET: 805 Third Avenue  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10022  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/963,825  
; FILING DATE:  
; CLASSIFICATION: 436  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/08/187,319  
; FILING DATE: 21-JAN-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Gogoris, Adda C  
; REGISTRATION NUMBER: 29,714  
; REFERENCE/DOCKET NUMBER: 4305/08701  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212-527-7700  
; TELEFAX: 212-753-6237  
; TELEX: 236687  
; INFORMATION FOR SEQ ID NO: 21:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1078 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; ORIGINAL SOURCE:  
; ORGANISM: Homo sapiens  
; IMMEDIATE SOURCE:  
; CLONE: COLLAGEN ALPHA 1 (III)  
US-08-963-825-21

Query Match 90.3%; Score 56; DB 3; Length 1078;  
Best Local Similarity 91.7%; Pred. No. 0.65;  
Matches 11; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 IAGIGGKAGGF 12  
| | | | | | | | | | | | | | | | | | | | | |  
Db 1058 IAGIGAEGKAGGF 1069

## RESULT 6

US-09-500-811-21  
; Sequence 21, Application US/09500811  
; Patent No. 6323314  
; GENERAL INFORMATION:  
; APPLICANT: Qvist, Per  
; APPLICANT: Bonde, Martin  
; TITLE OF INVENTION: A Method for Assaying Collagen Fragments  
; TITLE OF INVENTION: in Body Fluids, A Test Kit and Means for Carrying Out the  
; TITLE OF INVENTION: Method and Use of the Method to Diagnose the Presence of  
; TITLE OF INVENTION: Disorders Associated with the Metabolism of  
; NUMBER OF SEQUENCES: 21  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Darby & Darby PC  
; STREET: 805 Third Avenue  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10022  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/500,811  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/187,319  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Gogoris, Adda C  
; REGISTRATION NUMBER: 29,714  
; REFERENCE/DOCKET NUMBER: 4305/08701  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212-527-7700  
; TELEFAX: 212-753-6237  
; TELEX: 236687  
; INFORMATION FOR SEQ ID NO: 21:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1078 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; ORIGINAL SOURCE:  
; ORGANISM: Homo sapiens  
; IMMEDIATE SOURCE:  
; CLONE: COLLAGEN ALPHA 1 (III)  
US-09-500-811-21

Query Match 90.3%; Score 56; DB 4; Length 1078;  
Best Local Similarity 91.7%; Pred. No. 0.65;  
Matches 11; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 IAGIGGKAGGF 12  
| | | | | | | | | | | | | | | | | | | | | |  
Db 1058 IAGIGAEGKAGGF 1069

## RESULT 7

US-09-570-573-21  
; Sequence 21, Application US/09570573

; Patent No. 6342361  
; GENERAL INFORMATION:  
; APPLICANT: Qvist, Per  
; APPLICANT: Bonde, Martin  
; TITLE OF INVENTION: A Method for Assaying Collagen Fragments  
; TITLE OF INVENTION: in Body Fluids, A Test Kit and Means for Carrying Out the  
; TITLE OF INVENTION: Method and Use of the Method to Diagnose the Presence of  
; TITLE OF INVENTION: Disorders Associated with the Metabolism of  
; NUMBER OF SEQUENCES: 21  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Darby & Darby PC  
; STREET: 805 Third Avenue  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10022  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/570,573  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/187,319  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Gogoris, Adda C  
; REGISTRATION NUMBER: 29,714  
; REFERENCE/DOCKET NUMBER: 4305/08701  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212-527-7700  
; TELEFAX: 212-753-6237  
; TELEX: 236687  
; INFORMATION FOR SEQ ID NO: 21:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1078 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; ORIGINAL SOURCE:  
; ORGANISM: Homo sapiens  
; IMMEDIATE SOURCE:  
; CLONE: COLLAGEN ALPHA 1 (III)  
US-09-570-573-21

Query Match 90.3%; Score 56; DB 4; Length 1078;  
Best Local Similarity 91.7%; Pred. No. 0.65;  
Matches 11; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 IAGIGGKAGGF 12  
| | | | | | | | | | | | | | | | | | | | | |  
Db 1058 IAGIGAEGKAGGF 1069

## RESULT 8

US-09-548-608-21  
; Sequence 21, Application US/09548608  
; Patent No. 6355442

; GENERAL INFORMATION:  
; APPLICANT: Qvist, Per  
; APPLICANT: Bonde, Martin  
; TITLE OF INVENTION: A Method for Assaying Collagen Fragments  
; TITLE OF INVENTION: in Body Fluids, A Test Kit and Means for Carrying Out the  
; TITLE OF INVENTION: Method and Use of the Method to Diagnose the Presence of  
; TITLE OF INVENTION: Disorders Associated with the Metabolism of  
; NUMBER OF SEQUENCES: 21  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Darby & Darby PC  
; STREET: 805 Third Avenue  
; CITY: New York

```
; STATE: New York
; COUNTRY: USA
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/548,608
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/187,319
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Gogoris, Adda C
; REGISTRATION NUMBER: 29,714
; REFERENCE/DOCKET NUMBER: 4305/08701
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-527-7700
; TELEFAX: 212-753-6237
; TELEX: 236687
; INFORMATION FOR SEQ ID NO: 21:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1078 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; IMMEDIATE SOURCE:
; CLONE: COLLAGEN ALPHA 1 (III)
; US-09-548-608-21

Query Match 90.3%; Score 56; DB 4; Length 1078;
Best Local Similarity 91.7%; Pred. No. 0.65;
Matches 11; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 IAGIGGEKAGGF 12
DB 1058 IAGIGAEGKAGGF 1069

RESULT 9
US-08-278-774-20
; Sequence 20, Application US/08278774
; Patent No. 6653450
; GENERAL INFORMATION:
; APPLICANT: Berg, Richard A
; APPLICANT: Tomam, David P
; APPLICANT: Wallace, Donald
; TITLE OF INVENTION: MUTATED RECOMBINANT COLLAGENS
; NUMBER OF SEQUENCES: 33
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: COLLAGEN CORPORATION
; STREET: 2500 Faber Place
; CITY: Palo Alto
; STATE: California
; COUNTRY: USA
; ZIP: 94303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/278,774
; FILING DATE: 22-JUL-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Ralayko, Kathi L
; REGISTRATION NUMBER: 36,644
```

```
; REFERENCE/DOCKET NUMBER: 94-018
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 354-4642
; TELEFAX: (415) 354-4752
; TELEX:
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 15 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-278-774-20

Query Match 87.1%; Score 54; DB 4; Length 15;
Best Local Similarity 100.0%; Pred. No. 0.019;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 GIGGEKAGGF 12
DB 1 GIGGEKAGGF 10

RESULT 10
US-10-009-999A-41
; Sequence 41, Application US/10009999A
; Patent No. 6602980
; GENERAL INFORMATION:
; APPLICANT: Eyre, David R.
; TITLE OF INVENTION: SYNTHETIC PEPTIDES FOR COLLAGEN RESORPTION ASSAYS
; FILE REFERENCE: WROS-1-18220
; CURRENT APPLICATION NUMBER: US/10/009,999A
; CURRENT FILING DATE: 2003-01-22
; PRIOR APPLICATION NUMBER: PCT/US99/29357
; PRIOR FILING DATE: 1999-12-10
; PRIOR APPLICATION NUMBER: US 09/335,098
; PRIOR FILING DATE: 1999-06-17
; PRIOR APPLICATION NUMBER: US 60/141,574
; PRIOR FILING DATE: 1999-06-29
; PRIOR APPLICATION NUMBER: US 60/142,274
; PRIOR FILING DATE: 1999-07-02
; PRIOR APPLICATION NUMBER: US 60/142,675
; PRIOR FILING DATE: 1999-07-07
; PRIOR APPLICATION NUMBER: US 09/385,740
; PRIOR FILING DATE: 1999-08-30
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: Patentin version 3.2
; SEQ ID NO 41
; LENGTH: 9
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: MISC FEATURE
; OTHER INFORMATION: synthetic peptide corresponding to carboxy-terminal
; OTHER INFORMATION: telopeptide sequence of human type III collagen
; US-10-009-999A-41

Query Match 77.4%; Score 48; DB 4; Length 9;
Best Local Similarity 100.0%; Pred. No. 3e+05;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 IGGEKAGGF 12
DB 1 IGGEKAGGF 9

RESULT 11
5473052-20
; Patent No. 5473052
; APPLICANT: EYRE, DAVID R.
; TITLE OF INVENTION: ANTIGEN-BINDING FRAGMENTS OF AN ANTIBODY
; TO TYPE-I COLLAGEN AMINO-TERMINAL TELEPEPTIDE
; NUMBER OF SEQUENCES: 30
```

```

; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/221.705
; FILING DATE: 01-APR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 614.719
; FILING DATE: 21-NOV-1990
; APPLICATION NUMBER: 444.881
; FILING DATE: 01-DEC-1989
; APPLICATION NUMBER: 118.234
; FILING DATE: 06-NOV-1987
; SEQ ID NO:20:
; LENGTH: 11
; 5473052-20

Query Match      77.4%; Score 48; DB 6; Length 11;
Best Local Similarity 90.0%; Pred. No. 0.12;
Matches 9; Conservative 0; Mismatches 1; Indels 1; Gaps 0;

Qy 3 GIGGEKAGGF 12
Db 1 GIGGEXAGGF 10
      |||||
      |||||

RESULT 12
US-09-252-991A-30541
; Sequence 30541, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252.991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 30541
; LENGTH: 227
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
; US-09-252-991A-30541

Query Match      71.0%; Score 44; DB 4; Length 227;
Best Local Similarity 58.3%; Pred. No. 10;
Matches 7; Conservative 3; Mismatches 2; Indels 2; Gaps 0;

Qy 1 IAGIGGEKAGGF 12
Db 158 LAGAGGDRRGGF 169
      :|||
      :|||

RESULT 13
US-10-009-999A-40
; Sequence 40, Application US/10009999A
; Patent No. 6602980
; GENERAL INFORMATION:
; APPLICANT: Eyre, David R.
; TITLE OF INVENTION: SYNTHETIC PEPTIDES FOR COLLAGEN RESORPTION ASSAYS
; FILE REFERENCE: WR0S-1-18220
; CURRENT APPLICATION NUMBER: US/10/009.999A
; CURRENT FILING DATE: 2003-01-22
; PRIOR APPLICATION NUMBER: PCT/US99/29357
; PRIOR FILING DATE: 1999-12-10
; PRIOR APPLICATION NUMBER: US 09/335,098
; PRIOR FILING DATE: 1999-06-17
; PRIOR APPLICATION NUMBER: US 60/141,574
; PRIOR FILING DATE: 1999-06-29
; PRIOR APPLICATION NUMBER: US 60/142,274
; PRIOR FILING DATE: 1999-07-02
; PRIOR APPLICATION NUMBER: US 60/142,675

; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US 09/385,740
; FILING DATE: 1999-08-30
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 40
; LENGTH: 8
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: MISC FEATURE
; OTHER INFORMATION: synthetic peptide corresponding to carboxy-terminal
; OTHER INFORMATION: telopeptide sequence of human type III collagen
; US-10-009-999A-40

Query Match      67.7%; Score 42; DB 4; Length 8;
Best Local Similarity 100.0%; Pred. No. 3e+05;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 IGGEKAGG 11
Db 1 IGGEKAGG 8
      |||||
      |||||

RESULT 14
US-09-252-991A-28805
; Sequence 28805, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252.991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 28805
; LENGTH: 221
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
; US-09-252-991A-28805

Query Match      66.1%; Score 41; DB 4; Length 221;
Best Local Similarity 63.6%; Pred. No. 30;
Matches 7; Conservative 3; Mismatches 1; Indels 1; Gaps 0;

Qy 1 IAGIGGEKAGG 11
Db 186 IAALGGEQSGG 196
      |||
      |||

RESULT 15
US-09-252-991A-23643
; Sequence 23643, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252.991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 23643
; LENGTH: 503

```

: TYPE: PRT  
: ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-23643

Query Match 66.1%; Score 41; DB 4; Length 503;  
Best Local Similarity 70.0%; Pred. No. 67;  
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 AGIGEXKAGG 11  
||: ||: ||  
Db 481 AGLAGERAGG 490

Search completed: September 18, 2004, 04:32:23  
Job time : 19.7647 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 18, 2004, 04:29:56 ; Search time 56.9412 Seconds  
(without alignments)  
67.677 Million cell updates/sec

Title: US-10-615-959-43  
Perfect score: 62  
Sequence: 1 IAGIGGEKAGGF 12

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1342398 seqs, 321133274 residues

Total number of hits satisfying chosen parameters: 1342398

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications AA.\*

- 1: /cgn2\_6/ptodata/2/pubaa/US07\_PUBCOMB.pep.\*
- 2: /cgn2\_6/ptodata/2/pubaa/PCT\_NEW\_PUB.pep.\*
- 3: /cgn2\_6/ptodata/2/pubaa/US06\_NEW\_PUB.pep.\*
- 4: /cgn2\_6/ptodata/2/pubaa/US06\_PUBCOMB.pep.\*
- 5: /cgn2\_6/ptodata/2/pubaa/US07\_NEW\_PUB.pep.\*
- 6: /cgn2\_6/ptodata/2/pubaa/PCTUS\_PUBCOMB.pep.\*
- 7: /cgn2\_6/ptodata/2/pubaa/US08\_NEW\_PUB.pep.\*
- 8: /cgn2\_6/ptodata/2/pubaa/US08\_PUBCOMB.pep.\*
- 9: /cgn2\_6/ptodata/2/pubaa/US09A\_PUBCOMB.pep.\*
- 10: /cgn2\_6/ptodata/2/pubaa/US09B\_PUBCOMB.pep.\*
- 11: /cgn2\_6/ptodata/2/pubaa/US09C\_PUBCOMB.pep.\*
- 12: /cgn2\_6/ptodata/2/pubaa/US09\_NEW\_PUB.pep.\*
- 13: /cgn2\_6/ptodata/2/pubaa/US10A\_PUBCOMB.pep.\*
- 14: /cgn2\_6/ptodata/2/pubaa/US10B\_PUBCOMB.pep.\*
- 15: /cgn2\_6/ptodata/2/pubaa/US10C\_PUBCOMB.pep.\*
- 16: /cgn2\_6/ptodata/2/pubaa/US10\_NEW\_PUB.pep.\*
- 17: /cgn2\_6/ptodata/2/pubaa/US60\_NEW\_PUB.pep.\*
- 18: /cgn2\_6/ptodata/2/pubaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	62	100.0	12	12 US-10-615-959-43	Sequence 43, Appl
2	62	100.0	25	15 US-10-366-125-11	Sequence 11, Appl
3	62	100.0	262	15 US-10-366-125-19	Sequence 19, Appl
4	62	100.0	309	9 US-09-908-711-78	Sequence 78, Appl
5	62	100.0	1466	12 US-09-918-715-226	Sequence 226, App
6	62	100.0	1466	12 US-10-257-021-72	Sequence 72, Appl
7	62	100.0	1466	14 US-10-177-293-68	Sequence 68, Appl
8	62	100.0	1466	14 US-10-301-822-33	Sequence 33, Appl
9	62	100.0	1466	16 US-10-357-851-3	Sequence 3, Appl
10	62	100.0	1466	16 US-10-358-024-3	Sequence 3, Appl
11	62	100.0	1466	16 US-10-734-564-103	Sequence 103, App
12	61	98.4	1466	15 US-10-402-089-12	Sequence 12, Appl
13	61	98.4	1466	15 US-10-402-072A-12	Sequence 12, Appl
14	56	90.3	11	12 US-10-615-959-42	Sequence 42, Appl
15	56	90.3	1078	14 US-10-058-124-21	Sequence 21, Appl

16	55	88.7	1466	15	US-10-402-089-4	Sequence 4, Appli
17	55	88.7	1466	15	US-10-402-089-6	Sequence 6, Appli
18	55	88.7	1466	15	US-10-402-072A-4	Sequence 4, Appli
19	55	88.7	1466	15	US-10-402-072A-6	Sequence 6, Appli
20	48	77.4	9	12	US-10-615-959-41	Sequence 41, Appl
21	45	72.6	284	12	US-10-425-114-55247	Sequence 55247, A
22	43	69.4	225	16	US-10-437-963-106003	Sequence 106003,
23	42	67.7	80	12	US-10-615-959-40	Sequence 40, Appl
24	42	67.7	80	16	US-10-437-963-196331	Sequence 196331,
25	42	67.7	112	16	US-10-437-963-139908	Sequence 139908,
26	42	67.7	847	12	US-10-425-114-55135	Sequence 55135, A
27	41	66.1	49	16	US-10-437-963-163129	Sequence 163129,
28	41	66.1	105	16	US-10-767-701-37543	Sequence 37543, A
29	41	66.1	198	16	US-10-437-963-114393	Sequence 114393,
30	41	66.1	228	12	US-10-627-476-402	Sequence 402, App
31	41	66.1	238	12	US-10-425-114-37899	Sequence 37899, A
32	41	66.1	340	16	US-10-767-701-33026	Sequence 33026, A
33	41	66.1	691	12	US-10-282-122A-47568	Sequence 47568, A
34	40	64.5	105	16	US-10-767-701-36450	Sequence 36450, A
35	40	64.5	107	12	US-10-425-114-37950	Sequence 37950, A
36	40	64.5	284	16	US-10-437-963-128057	Sequence 128057,
37	40	64.5	300	12	US-10-425-114-47773	Sequence 47773, A
38	40	64.5	304	12	US-10-425-114-61720	Sequence 61720, A
39	40	64.5	438	10	US-09-988-067B-30	Sequence 30, Appl
40	40	64.5	438	12	US-10-335-977-6127	Sequence 6127, Ap
41	40	64.5	441	12	US-10-335-977-6128	Sequence 6128, Ap
42	40	64.5	467	12	US-10-425-114-70808	Sequence 70808, A
43	40	64.5	471	12	US-10-425-114-64617	Sequence 64617, A
44	40	64.5	479	12	US-10-282-122A-51430	Sequence 51430, A
45	40	64.5	696	16	US-10-437-963-189506	Sequence 189506,

## ALIGNMENTS

### RESULT 1

US-10-615-959-43  
; Sequence 43, Application US/10615959  
; Publication No. US200400048321A1  
; GENERAL INFORMATION:  
; APPLICANT: Eyre, David R.  
; TITLE OF INVENTION: SYNTHETIC PEPTIDES FOR COLLAGEN RESORPTION ASSAYS  
; FILE REFERENCE: WPOS-1-18220  
; CURRENT APPLICATION NUMBER: US/10/615,959  
; PRIOR FILING DATE: 2003-07-06  
; PRIOR APPLICATION NUMBER: US/10/009,999A  
; PRIOR FILING DATE: 2003-01-22  
; PRIOR APPLICATION NUMBER: PCT/US99/29357  
; PRIOR FILING DATE: 1999-12-10  
; PRIOR APPLICATION NUMBER: US 09/335,098  
; PRIOR FILING DATE: 1999-06-17  
; PRIOR APPLICATION NUMBER: US 60/141,574  
; PRIOR FILING DATE: 1999-06-29  
; PRIOR APPLICATION NUMBER: US 60/142,274  
; PRIOR FILING DATE: 1999-07-02  
; PRIOR APPLICATION NUMBER: US 60/142,675  
; PRIOR FILING DATE: 1999-07-07  
; PRIOR APPLICATION NUMBER: US 09/385,740  
; PRIOR FILING DATE: 1999-08-30  
; NUMBER OF SEQ ID NOS: 45  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 43  
; LENGTH: 12  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: MISC FEATURE  
; OTHER INFORMATION: Synthetic peptide corresponding to carboxy-terminal  
; OTHER INFORMATION: telopeptide sequence of human type III collagen  
US-10-615-959-43

Query Match 100.0%; Score 62; DB 12; Length 12;  
Best Local Similarity 100.0%; Pred. No. 0.0054;

Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IAGIGGEKAGGF 12  
|||||  
Db 1 IAGIGGEKAGGF 12

## RESULT 2

US-10-366-125-11  
; Sequence 11, Application US/10366125  
; Publication No. US20030228259A1  
; GENERAL INFORMATION:  
; APPLICANT: Hellerstein, Marc  
; TITLE OF INVENTION: MEASUREMENT OF BIOSYNTHESIS AND BREAKDOWN RATES OF  
; TITLE OF INVENTION: BIOLOGICAL MOLECULES THAT ARE INACCESSIBLE OR NOT  
; TITLE OF INVENTION: EASILY ACCESSIBLE TO DIRECT SAMPLING, NON-INVASIVELY,  
; TITLE OF INVENTION: BY LABEL INCORPORATION INTO METABOLIC DERIVATIVES AND  
; TITLE OF INVENTION: CATABOLITIC PRODUCTS  
; FILE REFERENCE: 416272003500  
; CURRENT APPLICATION NUMBER: US/10/366,125  
; CURRENT FILING DATE: 2003-02-12  
; PRIOR APPLICATION NUMBER: US 60/356,008  
; PRIOR FILING DATE: 2002-02-12  
; NUMBER OF SEQ ID NOS: 28  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 11  
; LENGTH: 25  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-366-125-11

Query Match 100.0%; Score 62; DB 15; Length 25;

Best Local Similarity 100.0%; Pred. No. 0.011;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 IAGIGGEKAGGF 12  
|||||  
Db 9 IAGIGGEKAGGF 20

## RESULT 3

US-10-366-125-19  
; Sequence 19, Application US/10366125  
; Publication No. US20030228259A1  
; GENERAL INFORMATION:  
; APPLICANT: Hellerstein, Marc  
; TITLE OF INVENTION: MEASUREMENT OF BIOSYNTHESIS AND BREAKDOWN RATES OF  
; TITLE OF INVENTION: BIOLOGICAL MOLECULES THAT ARE INACCESSIBLE OR NOT  
; TITLE OF INVENTION: EASILY ACCESSIBLE TO DIRECT SAMPLING, NON-INVASIVELY,  
; TITLE OF INVENTION: BY LABEL INCORPORATION INTO METABOLIC DERIVATIVES AND  
; TITLE OF INVENTION: CATABOLITIC PRODUCTS  
; FILE REFERENCE: 416272003500  
; CURRENT APPLICATION NUMBER: US/10/366,125  
; CURRENT FILING DATE: 2003-02-12  
; PRIOR APPLICATION NUMBER: US 60/356,008  
; PRIOR FILING DATE: 2002-02-12  
; NUMBER OF SEQ ID NOS: 28  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 19  
; LENGTH: 262  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-366-125-19

Query Match 100.0%; Score 62; DB 15; Length 262;

Best Local Similarity 100.0%; Pred. No. 0.12;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 IAGIGGEKAGGF 12  
|||||  
Db 1 IAGIGGEKAGGF 12

## RESULT 4

US-09-908-711-78  
; Sequence 78, Application US/09908711  
; Patent No. US20020045230A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: PA128  
; CURRENT APPLICATION NUMBER: US/09/908,711  
; CURRENT FILING DATE: 2001-07-20  
; PRIOR APPLICATION NUMBER: US01/01360  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: 09/764,867  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: US01/01344  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: 09/764,892  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: US01/01345  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: 09/764,888  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: US01/01329  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: 09/764,905  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: US01/01354  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: 09/764,891  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: US01/01339  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: 09/764,869  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: US01/01340  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: 09/764,874  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: US01/01334  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: 09/764,898  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: US01/01320  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: 09/764,853  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: US01/01349  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: 09/764,902  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: US01/01239  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: 09/764,870  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: US01/01348  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: 09/764,882  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: US01/01347  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: 09/764,896  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: US01/01307  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: 09/764,864  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: US01/01341  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: 09/764,856  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: US01/01336  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: 09/764,868

; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: US01/01312  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: 60/179,065  
; PRIOR FILING DATE: 2000-01-31  
; PRIOR APPLICATION NUMBER: 60/180,628  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: 60/209,467  
; PRIOR FILING DATE: 2000-06-07  
; NUMBER OF SEQ ID NOS: 167  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 78  
; LENGTH: 309

; TYPE: PRT  
; ORGANISM: Homo sapiens  
; NAME/KEY: SITE  
; LOCATION: (2)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE  
; LOCATION: (4)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
US-09-908-711-78

Query Match 100.0%; Score 62; DB 9; Length 309;  
Best Local Similarity 100.0%; Pred. No. 0.14;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 IAGIGGKAGGF 12  
||| ||||| |||||  
Db 48 IAGIGGKAGGF 59

## RESULT 5

US-09-918-715-226  
; Sequence 226, Application US/09918715  
; Publication No. US20030017157A1  
; GENERAL INFORMATION:  
; APPLICANT: Brad St. Croix  
; APPLICANT: Bert Vogelstein  
; APPLICANT: Kenneth Kinzler  
; TITLE OF INVENTION: ENDOTHELIAL CELL EXPRESSION PATTERNS  
; FILE REFERENCE: 1107.00134  
; CURRENT APPLICATION NUMBER: US/09/918,715  
; CURRENT FILING DATE: 2001-08-01  
; PRIOR APPLICATION NUMBER: 60/222,599  
; PRIOR FILING DATE: 2000-08-02  
; PRIOR APPLICATION NUMBER: 60/224,360  
; PRIOR FILING DATE: 2000-08-11  
; PRIOR APPLICATION NUMBER: 60/282,850  
; PRIOR FILING DATE: 2000-04-11  
; NUMBER OF SEQ ID NOS: 358  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 226  
; LENGTH: 1466  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-918-715-226

Query Match 100.0%; Score 62; DB 12; Length 1466;  
Best Local Similarity 100.0%; Pred. No. 0.67;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 IAGIGGKAGGF 12  
||| ||||| |||||  
Db 1205 IAGIGGKAGGF 1216

## RESULT 6

US-10-257-021-72  
; Sequence 72, Application US/10257021  
; Publication No. US20030211498A1  
; GENERAL INFORMATION:

; APPLICANT: Morin, Patrice J.  
; APPLICANT: Sherman-Baust, Cheryl A.  
; APPLICANT: Pizer, Ellen S.  
; APPLICANT: Hough, Colleen D.  
; TITLE OF INVENTION: TUMOR MARKERS IN OVARIAN CANCER  
; FILE REFERENCE: 14014.0369U2  
; CURRENT APPLICATION NUMBER: US/10/257,021  
; CURRENT FILING DATE: 2002-10-03  
; PRIOR APPLICATION NUMBER: PCT/US01/10947  
; PRIOR FILING DATE: 2001-04-03  
; PRIOR APPLICATION NUMBER: 60/194,336  
; PRIOR FILING DATE: 2000-04-03  
; NUMBER OF SEQ ID NOS: 147  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 72  
; LENGTH: 1466  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-257-021-72

Query Match 100.0%; Score 62; DB 12; Length 1466;  
Best Local Similarity 100.0%; Pred. No. 0.67;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 IAGIGGKAGGF 12  
||| ||||| |||||  
Db 1205 IAGIGGKAGGF 1216

## RESULT 7

US-10-177-293-68  
; Sequence 68, Application US/10177293  
; Publication No. US20030124128A1  
; GENERAL INFORMATION:  
; APPLICANT: Lillie, James  
; APPLICANT: Glat, Karen  
; APPLICANT: Zhao, Xumei  
; APPLICANT: Gannavarpu, Manjula  
; APPLICANT: Kamatkar, Shubhangi  
; APPLICANT: Mertens, Maureen  
; APPLICANT: Myer, Vic  
; APPLICANT: Wang, Youzhen  
; APPLICANT: Xu, Yongyao  
; APPLICANT: Hoersch, Sebastian  
; APPLICANT: Morahan, John  
; APPLICANT: Meyers, Rachel E.  
; APPLICANT: Bast Jr., Robert C.  
; APPLICANT: Hortobagyi, Gabriel N.  
; APPLICANT: Pusztai, Lajos  
; APPLICANT: Meric, Funda  
; APPLICANT: Sahin, Aysegul  
; APPLICANT: Mills, Gordon B.  
; TITLE OF INVENTION: COMPOSITIONS, KITS, AND METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY OF BREAST CANCER  
; FILE REFERENCE: MRI-038  
; CURRENT APPLICATION NUMBER: US/10/177,293  
; CURRENT FILING DATE: 2002-06-21  
; PRIOR APPLICATION NUMBER: US 60/299,887  
; PRIOR FILING DATE: 2001-06-21  
; PRIOR APPLICATION NUMBER: US 60/301,572  
; PRIOR FILING DATE: 2001-06-27  
; PRIOR APPLICATION NUMBER: US 60/306,501  
; PRIOR FILING DATE: 2001-07-18  
; PRIOR APPLICATION NUMBER: US 60/325,002  
; PRIOR FILING DATE: 2001-09-25  
; PRIOR APPLICATION NUMBER: US 60/362,585  
; PRIOR FILING DATE: 2002-03-05  
; PRIOR APPLICATION NUMBER: US 60/xxx,xxx  
; PRIOR FILING DATE: 2002-05-14  
; NUMBER OF SEQ ID NOS: 506  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 68  
; LENGTH: 1466

```
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-177-293-68

Query Match      100.0%; Score 62; DB 14; Length 1466;
Best Local Similarity 100.0%; Pred. No. 0.67;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 IAGIGGEKAGGF 12
Db      1205 IAGIGGEKAGGF 1216

RESULT 8
US-10-301-822-33
; Sequence 33, Application US/10301822
; Publication No. US20030148410A1
; GENERAL INFORMATION:
; APPLICANT: Millennium Pharmaceuticals, Inc.
; APPLICANT: Berger, Allison
; APPLICANT: Guillemette, Tracy L.
; APPLICANT: Kamatkar, Shubhangi
; APPLICANT: Schlegel, Robert
; APPLICANT: Monahan, John E.
; APPLICANT: Thibodeau, Stephen N.
; APPLICANT: Burgart, Lawrence J.
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND
; TITLE OF INVENTION: METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF COLON CANCER
; FILE REFERENCE: MPM01-029P2RNM
; CURRENT APPLICATION NUMBER: US/10/301,822
; CURRENT FILING DATE: 2002-11-21
; PRIOR APPLICATION NUMBER: US 60/339,971
; PRIOR FILING DATE: 2001-12-10
; PRIOR APPLICATION NUMBER: US 60/361,978
; PRIOR FILING DATE: 2002-03-05
; PRIOR APPLICATION NUMBER: US 60/381,988
; PRIOR FILING DATE: 2002-05-20
; NUMBER OF SEQ ID NOS: 228
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 33
; LENGTH: 1466
; TYPE: PRT
; ORGANISM: Homo Sapiens
US-10-301-822-33

Query Match      100.0%; Score 62; DB 14; Length 1466;
Best Local Similarity 100.0%; Pred. No. 0.67;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 IAGIGGEKAGGF 12
Db      1205 IAGIGGEKAGGF 1216

RESULT 9
US-10-357-851-3
; Sequence 3, Application US/10357851
; Publication No. US20040151731A1
; GENERAL INFORMATION:
; APPLICANT: Jicha, Douglas L.
; TITLE OF INVENTION: Method and Compositions Involving
; TITLE OF INVENTION: T-Lymphocyte Reactivity with Collagen in Aortic Tissue of
; TITLE OF INVENTION: Abdominal Aortic Aneurysm Patients
; FILE REFERENCE: 13376US
; CURRENT APPLICATION NUMBER: US/10/357,851
; CURRENT FILING DATE: 2003-02-04
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 1466
; TYPE: PRT
; ORGANISM: Homo Sapiens

US-10-357-851-3
; Sequence 3, Application US/10358024
; Publication No. US20040151732A1
; GENERAL INFORMATION:
; APPLICANT: Jicha, Douglas L.
; APPLICANT: Pelsue, Stephen
; TITLE OF INVENTION: Methods and Compositions Involving Blood
; TITLE OF INVENTION: T-Lymphocytes Reactivity with Collagen
; FILE REFERENCE: 13436US
; CURRENT APPLICATION NUMBER: US/10/358,024
; CURRENT FILING DATE: 2003-02-04
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 1466
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-358-024-3

Query Match      100.0%; Score 62; DB 16; Length 1466;
Best Local Similarity 100.0%; Pred. No. 0.67;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 IAGIGGEKAGGF 12
Db      1205 IAGIGGEKAGGF 1216

RESULT 10
US-10-358-024-3
; Sequence 3, Application US/10358024
; Publication No. US20040151732A1
; GENERAL INFORMATION:
; APPLICANT: Jicha, Douglas L.
; APPLICANT: Pelsue, Stephen
; TITLE OF INVENTION: Methods and Compositions Involving Blood
; TITLE OF INVENTION: T-Lymphocytes Reactivity with Collagen
; FILE REFERENCE: 13436US
; CURRENT APPLICATION NUMBER: US/10/358,024
; CURRENT FILING DATE: 2003-02-04
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 1466
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-358-024-3

Query Match      100.0%; Score 62; DB 16; Length 1466;
Best Local Similarity 100.0%; Pred. No. 0.67;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 IAGIGGEKAGGF 12
Db      1205 IAGIGGEKAGGF 1216

RESULT 11
US-10-734-564-103
; Sequence 103, Application US/10734564
; Publication No. US20040157278A1
; GENERAL INFORMATION:
; APPLICANT: Christopher C Burgess et al
; TITLE OF INVENTION: Detection Methods Using TIMP1
; FILE REFERENCE: 1657/2012
; CURRENT APPLICATION NUMBER: US/10/734,564
; CURRENT FILING DATE: 2003-12-12
; NUMBER OF SEQ ID NOS: 138
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 103
; LENGTH: 1466
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-734-564-103

Query Match      100.0%; Score 62; DB 16; Length 1466;
Best Local Similarity 100.0%; Pred. No. 0.67;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 IAGIGGEKAGGF 12
Db      1205 IAGIGGEKAGGF 1216

RESULT 12
US-10-402-089-12
; Sequence 12, Application US/10402089
; Publication No. US20040005663A1
; GENERAL INFORMATION:
; APPLICANT: Bell, Marcum P.
```

; APPLICANT: Neff, Thomas B.  
; APPLICANT: Polarek, James W.  
; APPLICANT: Seeley, Todd W.  
; TITLE OF INVENTION: PORCINE COLLAGENS AND GELATINS  
; FILE REFERENCE: FP0402.3 CON  
; CURRENT APPLICATION NUMBER: US/10/402,089  
; CURRENT FILING DATE: 2003-03-26  
; PRIOR APPLICATION NUMBER: US 09/709,700  
; PRIOR FILING DATE: 2000-11-10  
; NUMBER OF SEQ ID NOS: 72  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 12  
; LENGTH: 1466  
; TYPE: PRT  
; ORGANISM: Sus scrofa  
US-10-402-089-12

Query Match 98.4%; Score 61; DB 15; Length 1466;  
Best Local Similarity 91.7%; Pred. No. 0.94;  
Matches 11; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 IAGIGGEKAGGF 12  
Db 1205 IAGVGEKAGGF 1216

RESULT 13  
US-10-402-072A-12  
; Sequence 12, Application US/10402072A  
; Publication No. US20040018592A1  
; GENERAL INFORMATION:  
; APPLICANT: Bell, Marcum P.  
; APPLICANT: Neff, Thomas B.  
; APPLICANT: Polarek, James W.  
; APPLICANT: Seeley, Todd W.  
; TITLE OF INVENTION: BOVINE COLLAGENS AND GELATINS  
; FILE REFERENCE: FP0402.2 CON  
; CURRENT APPLICATION NUMBER: US/10/402,072A  
; CURRENT FILING DATE: 2003-03-26  
; PRIOR APPLICATION NUMBER: US 09/709,700  
; PRIOR FILING DATE: 2000-11-10  
; NUMBER OF SEQ ID NOS: 72  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 12  
; LENGTH: 1466  
; TYPE: PRT  
; ORGANISM: Sus scrofa  
US-10-402-072A-12

Query Match 98.4%; Score 61; DB 15; Length 1466;  
Best Local Similarity 91.7%; Pred. No. 0.94;  
Matches 11; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 IAGIGGEKAGGF 12  
Db 1205 IAGVGEKAGGF 1216

RESULT 14  
US-10-615-959-42  
; Sequence 42, Application US/10615959  
; Publication No. US20040048321A1  
; GENERAL INFORMATION:  
; APPLICANT: Eyre, David R.  
; TITLE OF INVENTION: SYNTHETIC PEPTIDES FOR COLLAGEN RESORPTION ASSAYS  
; FILE REFERENCE: WR0S-1-18220  
; CURRENT APPLICATION NUMBER: US/10/615,959  
; CURRENT FILING DATE: 2003-07-08  
; PRIOR APPLICATION NUMBER: US/10/009,999A  
; PRIOR FILING DATE: 2003-01-22  
; PRIOR APPLICATION NUMBER: PCT/US99/29357  
; PRIOR FILING DATE: 1999-12-10  
; PRIOR APPLICATION NUMBER: US 09/335,098

; PRIOR FILING DATE: 1999-06-17  
; PRIOR APPLICATION NUMBER: US 60/142,574  
; PRIOR FILING DATE: 1999-06-29  
; PRIOR APPLICATION NUMBER: US 60/142,274  
; PRIOR FILING DATE: 1999-07-02  
; PRIOR APPLICATION NUMBER: US 60/142,675  
; PRIOR FILING DATE: 1999-07-07  
; PRIOR APPLICATION NUMBER: US 09/385,740  
; PRIOR FILING DATE: 1999-08-30  
; NUMBER OF SEQ ID NOS: 45  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 42  
; LENGTH: 11  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: MISC FEATURE  
; OTHER INFORMATION: synthetic peptide corresponding to carboxy-terminal  
; OTHER INFORMATION: telopeptide sequence of human type III collagen  
US-10-615-959-42

Query Match 90.3%; Score 56; DB 12; Length 11;  
Best Local Similarity 100.0%; Pred. No. 0.041;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 IAGIGGEKAGG 11  
Db 1 IAGIGGEKAGG 11

RESULT 15  
US-10-058-124-21  
; Sequence 21, Application US/10058124  
; Publication No. US20030119058A1  
; GENERAL INFORMATION:  
; APPLICANT: Oviatt, Per  
; TITLE OF INVENTION: A Method for Assaying Collagen Fragments  
; in Body Fluids, A Test Kit and Means for Carrying Out the  
; Method and Use of the Method to Diagnose the Presence of  
; Disorders Associated with the Metabolism of  
; NUMBER OF SEQUENCES: 21  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Darby & Darby PC  
; STREET: 805 Third Avenue  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10022  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/10/058,124  
; FILING DATE: 29-Jan-2002  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 09/570,573  
; FILING DATE: 2002-MAY-12  
; APPLICATION NUMBER: 08/187,319  
; FILING DATE: <Unknown>  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Gogoris, Adda C  
; REGISTRATION NUMBER: 29,714  
; REFERENCE/DOCKET NUMBER: 4305/08701  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212-527-7700  
; TELEFAX: 212-753-6237  
; TELEX: 236687  
; INFORMATION FOR SEQ ID NO: 21:  
; SEQUENCE CHARACTERISTICS:

```

; LENGTH: 1078 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; IMMEDIATE SOURCE:
; CLONE: COLLAGEN ALPHA 1 (III)
; SEQUENCE DESCRIPTION: SEQ ID NO: 21:
US-10-058-124-21

Query Match          90.3%; Score 56; DB 14; Length 1078;
Best Local Similarity 91.7%; Pred. No. 4;
Matches 11; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 IAGIGGEKAGGF 12
        ||||| |||||
Db      1058 IAGIGAEEKAGGF 1069

```

Search completed: September 18, 2004, 04:56:14  
Job time : 57.9412 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 18, 2004, 04:02:05 ; Search time 14.5882 Seconds  
(without alignments)  
79.125 Million cell updates/sec

Title: US-10-615-959-43

Perfect score: 62

Sequence: 1 IAGIGGKAGGF 12

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR 78.\*

1: pir1.\*

2: pir2.\*

3: pir3.\*

4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	62	100.0	1466	1 CGHU7L	collagen alpha 1(I)
2	58	93.5	636	2 S41067	collagen alpha 1(I)
3	58	93.5	1464	2 S59856	collagen alpha 1(I)
4	44	71.0	161	2 S14998	heat shock protein
5	44	71.0	487	2 F70863	probable folypoly
6	42	67.7	209	2 AE2981	arylesterase [limpo
7	42	67.7	209	2 C98302	arylesterase (AF04
8	41	66.1	383	2 G72777	probable S2P metal
9	41	66.1	1332	2 F69732	PBSX prophage ORF
10	41	66.1	1852	2 A37860	calcium channel pr
11	40	64.5	270	2 H69074	formylmethanofuran
12	40	64.5	270	2 S57457	formylmethanofuran
13	40	64.5	298	2 E95286	hypothetical prote
14	40	64.5	438	2 C64654	conserved hypothet
15	40	64.5	438	2 H71943	hypothetical prote
16	40	64.5	548	2 T51035	hypothetical prote
17	39	62.9	87	2 T14302	glycine-rich cell
18	39	62.9	163	2 T47394	hypothetical prote
19	39	62.9	175	2 T27543	hypothetical prote
20	39	62.9	399	2 I49754	homeobox protein -
21	39	62.9	605	2 H95240	conserved hypothet
22	39	62.9	605	2 A92905	hypothetical prote
23	39	62.9	622	2 AF0169	probable exported
24	39	62.9	636	2 T37843	probable beta-gluc
25	39	62.9	846	2 H70599	hypothetical prote
26	39	62.9	1559	2 T30535	calcium channel al
27	38	61.3	151	2 T08002	glycine-rich prote
28	38	61.3	205	2 F84274	halocyanin precurs
29	38	61.3	302	2 G70691	hypothetical prote

#### ALIGNMENTS

##### RESULT 1

CGHU7L

collagen alpha 1(III) chain precursor - human

N/Alternate names: procollagen alpha 1(III) chain

C/Species: Homo sapiens (man)

C/Date: 24-Apr-1984 #sequence revision 01-Sep-1995 #text change 21-Jul-2000

C/Accession: S05272; S04642; PE0011; S01726; S04887; A90399; A94562; I51868; S59511; A904

R/Prockop, D.J.

submitted to the EMBL Data Library, February 1989

A/Reference number: S05272

A/Accession: S05272

A/Status: preliminary

A/Molecule type: mRNA

A/Residues: 1-1240, 'V', 1242-1466 <PRC>

A/Cross-references: EMBL:X14420; NID:G30057; PIDN:CAA32583.1; PID:G30058

R/Ala-Kokko, L.; Kontusaari, S.; Baldwin, C.T.; Kuivaniemi, H.; Prockop, D.J.

Biochem. J. 260, 509-516, 1989

A/Title: Structure of cDNA clones coding for the entire prepro-alpha(III) chain of human

erences.

A/Reference number: S04642; MUID:89350838; PMID:2764886

A/Accession: S04642

A/Molecule type: mRNA

A/Residues: 1-1196 <ALA>

A/Cross-references: EMBL:X14420; NID:G30057; PIDN:CAA32583.1; PID:G30058

A/Note: the complete sequence is not shown

R/Benson-Chanda, V.; Su, M.W.; Weil, D.; Chu, M.L.; Ramirez, F.

Gene 78, 255-265, 1989

A/Title: Cloning and analysis of the 5' portion of the human type-III procollagen gene (C

A/Reference number: PE0011; MUID:89378752; PMID:2777083

A/Accession: PE0011

A/Molecule type: DNA

A/Residues: 1-176 <BEN>

A/Cross-references: GB:M26939; NID:G180813; PIDN:AAA52040.1; PID:G180814

R/Toman, P.D.; Ricca, G.A.; de Crombrughe, B.

Nucleic Acids Res. 16, 7201, 1988

A/Title: Nucleotide sequence of a cDNA coding for the amino-terminal region of human pref

A/Reference number: S01726; MUID:89303360; PMID:3405773

A/Accession: S01726

A/Molecule type: mRNA

A/Residues: 1-170 <TOM>

A/Cross-references: EMBL:X07240; NID:G30060; PIDN:CAA30229.1; PID:G30061

A/Note: the authors translated the codon CAG for residue 134 as His

R/Janczko, R.A.; Ramirez, F.

Nucleic Acids Res. 17, 6742, 1989

A/Title: Nucleotide and amino acid sequences of the entire human alpha-1 (III) collagen.

A/Reference number: S04887; MUID:89386015; PMID:2780304

A/Accession: S04887

A/Molecule type: mRNA

A/Residues: 149-163, 'G', 164-240, 'D', 242-471, 'D', 473-487, 'L', 489, 'S', 491-613, 'Y', 615-634, 'A'

A/Cross-references: EMBL:X15332; NID:G29545; PIDN:CAA33387.1; PID:G930045

A/Note: the authors' translation of residues 905-932 is inconsistent with the nucleotide

R/Seyer, J.M.; Kang, A.H.

circumsporozoite p  
leukosialin precu  
protein F2iH11.3 [  
leucine aminopepti  
hypothetical glyci  
RNA/ssDNA-binding  
hypothetical prote  
hypothetical prote  
hypothetical prote  
phage lambda-relat  
hypothetical prote  
coat protein - pas  
hypothetical prote  
hypothetical prote  
probable ATP-bind

Biochemistry 16, 1158-1164, 1977  
A;Title: Covalent structure of collagen: amino acid sequence of cyanogen bromide peptide  
A;Reference number: A90399; UID:77134724; PMID:557335  
A;Accession: A90399  
A;Molecule type: protein  
A;Residues: 'Y', 169-225, 229-232, 'P', 234-292, 'D', 294, 'S', 296-398 <SEY1>  
A;Experimental source: liver  
A;Note: sequence corrected by A94562; attachment of 2-O-alpha-D-glucosyl-O-beta-D-galactose  
R;Seyer, J.M.  
submitted to the Atlas, December 1977  
A;Reference number: A94562  
A;Accession: A94562  
A;Molecule type: protein  
A;Residues: 'Y', 169-225, 229-277, 'A', 279-292, 'D', 294, 'S', 296-398 <SEY2>  
A;Experimental source: liver  
A;Note: author submitted corrections to A90399  
R;Milewicz, D.M.; Witz, A.M.; Smith, A.C.; Manchester, D.K.; Waldstein, G.; Byers, P.H.  
Am. J. Hum. Genet. 53, 62-70, 1993  
A;Title: Parental somatic and germ-line mosaicism for a multiexon deletion with unusual  
splicing.  
A;Reference number: I51868; UID:93304430; PMID:8317500  
A;Accession: I51868  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 186-194 <MIL>  
A;Cross-references: GB:S62925; NID:G386425; PIDN:AAAD13937.1; PID:G4261637  
R;Chiodo, A.A.; Sillence, D.O.; Cole, W.G.; Bateman, J.F.  
Biochem. J. 311, 939-943, 1995  
A;Title: Abnormal type III collagen produced by an exon-17-skipping mutation of the COL3A1  
A;Reference number: S59511; UID:96067614; PMID:7487954  
A;Accession: S59511  
A;Molecule type: mRNA  
A;Residues: 302-423 <CHI>  
A;Cross-references: GB:S79877; NID:G1195576; PIDN:AAB35615.1; PID:G1195577  
R;Seyer, J.M.; Kang, A.H.  
Biochemistry 17, 3404-3411, 1978  
A;Title: Covalent structure of collagen: amino acid sequence of five consecutive CNBr pe  
A;Reference number: A90414; UID:79000343; PMID:687591  
A;Accession: A90414  
A;Molecule type: protein  
A;Residues: 399-675, 'N', 677-727 <SEY3>  
A;Experimental source: liver  
R;Lee, B.; Vitale, E.; Superti-Furga, A.; Steinmann, B.; Ramirez, F.  
J. Biol. Chem. 266, 5256-5259, 1991  
A;Title: G to T transversion at position +5 of a splice donor site causes skipping of th  
A;Reference number: I55349; UID:91161621; PMID:1672129  
A;Accession: I55349  
A;Status: translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 537-605 <LEE>  
A;Cross-references: GB:M59312; NID:G180815; PIDN:AAA52041.1; PID:G180816  
R;Seyer, J.M.; Mainardi, C.; Kang, A.H.  
Biochemistry 19, 1583-1589, 1980  
A;Title: Covalent structure of collagen: amino acid sequence of alpha1 (III)-CB5 from ty  
A;Reference number: A90438; UID:80198282; PMID:6246925  
A;Accession: A90438  
A;Molecule type: protein  
A;Residues: 728-895, 'A', 897-964 <SEY4>  
A;Experimental source: liver  
R;Cole, W.G.; Chiodo, A.A.; Lamande, S.R.; Janeczko, R.; Ramirez, F.; Dahl, H.H.M.; Char  
J. Biol. Chem. 265, 17070-17077, 1990  
A;Title: A base substitution at a splice site in the COL3A1 gene causes exon skipping an  
A;Reference number: A38303; UID:91009133; PMID:2145268  
A;Accession: A38303  
A;Molecule type: mRNA  
A;Residues: 861-1015 <COL>  
A;Cross-references: GB:J05617; GB:M55603; GB:M59227; NID:G180878; PIDN:AAB59383.1; PID:G  
A;Note: a mutant sequence with 942-977 spliced out from a patient with Ehlers-Danlos syn  
R;Mankoo, B.S.; Dalglish, R.  
Nucleic Acids Res. 16, 2337, 1988  
A;Title: Human pro alpha1(III) collagen: cDNA sequence for the 3' end.  
A;Reference number: S02119; UID:88189827; PMID:3357782  
A;Accession: S02119

A;Status: translation not shown  
A;Molecule type: mRNA  
A;Residues: 950-1018, 'Y', 1020-1183, 'S', 1185-1466 <MAN>  
A;Cross-references: EMBL:X06700; NID:G30053; PIDN:CAA29886.1; PID:G30054  
R;Seyer, J.M.; Kang, A.H.  
Biochemistry 20, 2621-2627, 1981  
A;Title: Covalent structure of collagen: amino acid sequence of alpha1 (III)-CB9 from ty  
A;Reference number: A90446; UID:81208139; PMID:7016180  
A;Accession: A90446  
A;Molecule type: protein  
A;Residues: 965-979, 'A', 981-984, 'PS', 987, 'QN', 990-1096, 'P', 1098-1152, 'AT', 1155, 'S', 1157-1  
A;Experimental source: liver  
R;Loidl, H.R.; Brinker, J.M.; May, M.; Pihlajaniemi, T.; Morrow, S.; Rosenbloom, J.; Myer  
Nucleic Acids Res. 12, 9383-9394, 1984  
A;Title: Molecular cloning and carboxyl-propeptide analysis of human type III procollagen  
A;Reference number: A93551; UID:85087944; PMID:6096827  
A;Accession: A93551  
A;Molecule type: mRNA  
A;Residues: 1065-1155, 'P', 1157-1466 <LOI>  
A;Cross-references: EMBL:X01655; EMBL:X01742; NID:G29584; PIDN:CAA25821.1  
R;Miskulin, M.; Dalglish, R.; Kluge-Becker, B.; Renard, S.L.; Tolstoshev, P.; Brant  
Biochemistry 25, 1408-1413, 1986  
A;Title: Human type III collagen gene expression is coordinately modulated with the type  
A;Reference number: I52393; UID:86187804; PMID:3754462  
A;Accession: I52393  
A;Status: translated from GB/EMBL/DBJ  
A;Molecule type: mRNA  
A;Residues: 1161-1200 <MIS>  
A;Cross-references: GB:M13146; NID:G180415; PIDN:AAA52003.1; PID:G180416  
R;Emanuel, B.S.; Cannizzaro, L.A.; Seyer, J.M.; Myers, J.C.  
Proc. Natl. Acad. Sci. U.S.A. 82, 3385-3389, 1985  
A;Title: Human alpha 1(III) and alpha 2(V) procollagen genes are located on the long arm  
A;Reference number: I59025; UID:85216505; PMID:3858826  
A;Accession: I59025  
A;Status: translated from GB/EMBL/DBJ  
A;Molecule type: mRNA  
A;Residues: 1165-1196 <EMA>  
A;Cross-references: GB:M11134; NID:G180417; PIDN:AAA52004.1; PID:G180418  
R;Chu, M.L.; Weil, D.; de Wet, W.; Bernard, M.; Sippola, M.; Ramirez, F.  
J. Biol. Chem. 260, 4357-4363, 1985  
A;Title: Isolation of cDNA and genomic clones encoding human pro-alpha1(III) collagen. P  
A;Reference number: A92516; UID:85157600; PMID:2579949  
A;Accession: A92516  
A;Molecule type: DNA  
A;Residues: 1176-1240, 'V', 1242-1356, 'P', 1358-1466 <CHU>  
A;Cross-references: GB:M10615; GB:M10793; GB:M10795; GB:M10796; GB:M10797; GB:M10798  
A;Experimental source: liver  
A;Note: the authors translated the codon TTC for residue 1057 as Tyr; the codons given fo  
ation  
C;Comment: Prolines and lysines at the third position of the tripeptide repeating unit (C  
3-hydroxylated. About 15% of the lysines are 5-hydroxylated and some are subsequently O  
C;Genetics:  
A;Gene: GDB:COL3A1  
A;Cross-references: GDB:118729; OMIM:120180  
A;Map position: 2q31-2q31  
A;Introns: 27/1; 94/3; 111/3; 149/3; 176/3; 554/3; 587/3; 1175/3; 1275/3; 1337/3; 1418/3  
A;Note: the list of introns is incomplete; defects in this gene can result in Ehlers-Danl  
C;Complex: type III collagen is a homotrimer of monomers initially linked by disulfide br  
er of their length, is formed with desmosine cross-links made from lysine and allylsine re  
C;Function:  
A;Description: structural component of extracellular fibrous polymer that maintains inter  
C;Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;  
C;Keywords: coiled coil; Ehlers-Danlos syndrome; extracellular matrix; glycoprotein; hyd  
F;1-23/Domain: signal sequence #status predicted <SIG>  
F;24-153/Domain: amino-terminal propeptide #status predicted <PRO>  
F;31-91/Domain: von Willebrand factor type C repeat homology <VWC>  
F;154-1221/Product: collagen alpha 1(III) chain #status predicted <MAT>  
F;154-167/Region: amino-terminal nonhelical telopeptide  
F;168-1196/Region: helical  
F;1091-1093/Region: cell attachment (R-G-D) motif  
F;1197-1221/Region: carboxyl-terminal nonhelical telopeptide  
F;1222-1466/Domain: carboxyl-terminal propeptide #status predicted <CPR>  
F;1238-1466/Domain: fibrillar collagen carboxyl-terminal homology <FCC>



F;24/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status predicted  
 F;153-154/Cleavage site: Pro-Gln (procollagen N-endopeptidase) #status predicted  
 F;154/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status predicted  
 F;161,1212/Modified site: allysine (Lys) #status predicted  
 F;263,284,860,977,1106/Modified site: 5-hydroxylysine (Lys) #status experimental  
 F;263/Banding site: carbohydrate (Lys) (covalent) #status experimental  
 F;584,1094/Modified site: 5-hydroxylysine (Lys) (partial) #status experimental  
 F;948-949/Cleavage site: Gly-Ile (collagenase) #status experimental  
 F;1106/Binding site: carbohydrate (Lys) (covalent) #status predicted

Query Match 100.0%; Score 62; DB 1; Length 1466;  
 Best Local Similarity 100.0%; Pred. No. 0.053;  
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 IAGIGGEKAGGF 12  
 |||||  
 Db 1205 IAGIGGEKAGGF 1216

RESULT 2  
 S41067  
 collagen alpha 1(III) chain - rat  
 C;Species: Rattus norvegicus (Norway rat)  
 C;Date: 13-Jan-1995 #sequence revision 13-Jan-1995 #text\_change 13-Aug-1999  
 C;Accession: S41067; A29905; S31924  
 R;Glumoff, V.; Maekelae, J.K.; Vuorio, E.  
 Biochim. Biophys. Acta 1217, 41-48, 1994  
 A;Title: Cloning of cDNA for rat pro alpha-1(III) collagen mRNA. Different expression pa  
 A;Reference number: S41067; MUID:94114571; PMID:8286415  
 A;Accession: S41067  
 A;Status: preliminary  
 A;Molecule type: mRNA  
 A;Residues: 1-636 <GLU>  
 A;Cross-references: EMBL:X70369; NID:G57915; PIDN:CAA49832.1; PID:G57916  
 R;Frankel, F.R.; Hau, C.Y.J.; Myers, J.C.; Lin, B.; Lyttle, C.R.; Komm, B.; Mohn, K.  
 DNA 7, 347-354, 1988  
 A;Title: Regulation of alpha-2 (I), alpha-1 (III), and alpha-2 (V) collagen mRNAs by estr  
 A;Reference number: A29905; MUID:88296083; PMID:2456904  
 A;Accession: A29905  
 A;Molecule type: mRNA  
 A;Residues: 308-482 <FRA>  
 A;Cross-references: GB:M21354; NID:G203500; PIDN:AAA40942.1; PID:G203501  
 R;Glumoff, V.; Maekelae, J.K.; Vuorio, E.  
 submitted to the EMBL Data Library, February 1993  
 A;Reference number: S31924  
 A;Accession: S31924  
 A;Status: preliminary  
 A;Molecule type: mRNA  
 A;Residues: 2-636 <GL2>  
 A;Cross-references: EMBL:X70369  
 C;Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;  
 C;Keywords: coiled coil; extracellular matrix; glycoprotein; trimer; triple helix  
 F;408-636/Domain: fibrillar collagen carboxyl-terminal homology <FCC>

Query Match 93.5%; Score 58; DB 2; Length 636;  
 Best Local Similarity 83.3%; Pred. No. 0.11;  
 Matches 10; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 IAGIGGEKAGGF 12  
 |||||  
 Db 375 IAGVGGEKSGGF 386

RESULT 3  
 S59856  
 collagen alpha 1(III) chain precursor - mouse  
 C;Species: Mus musculus (house mouse)  
 C;Date: 10-Apr-1996 #sequence revision 19-Apr-1996 #text\_change 13-Aug-1999  
 C;Accession: S59856; S62120; S16373  
 R;Toman, P.D.; de Crombrughe, B.  
 Gene 147, 161-168, 1994  
 A;Title: The mouse type-III procollagen-encoding gene: genomic cloning and complete DNA  
 A;Reference number: S59856; MUID:95011609; PMID:7926795

A;Accession: S59856  
 A;Molecule type: DNA  
 A;Residues: 1-1464 <TOM>  
 A;Cross-references: EMBL:X52046  
 R;Toman, D.  
 submitted to the EMBL Data Library, November 1994  
 A;Reference number: S62120  
 A;Accession: S62120  
 A;Molecule type: DNA  
 A;Residues: 1-866, 'G', 868-1464 <TOA>  
 A;Cross-references: EMBL:X52046; NID:G575321; PIDN:CAA36279.1; PID:G575322  
 R;Metsaeranta, M.; Toman, D.; de Crombrughe, B.; Vuorio, E.  
 Biochim. Biophys. Acta 1089, 241-243, 1991  
 A;Title: Specific hybridization probes for mouse type I, II, III and IX collagen mRNAs.  
 A;Reference number: S16176; MUID:91274355; PMID:2054384  
 A;Accession: S16373  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1442-1464 <MET>  
 A;Cross-references: EMBL:X57983; NID:G50476; PIDN:CAA41048.1; PID:G50477  
 C;Genetics:  
 A;Introns: 29/1; 95/3; 112/3; 150/3; 175/3; 193/3; 211/3; 229/3; 247/3; 265/3; 283/3; 291/3; 58/3; 673/3; 706/3; 742/3; 760/3; 778/3; 796/3; 814/3; 850/3; 868/3; 886/3; 940/3; 976/3;  
 C;Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;  
 C;Keywords: coiled coil; extracellular matrix  
 F;1-24/Domain: signal sequence #status predicted <SIG>  
 F;25-154/Domain: propeptide #status predicted <PRO>  
 F;32-92/Domain: von Willbrand factor type C repeat homology <VWC>  
 F;155-1464/Product: collagen alpha 1(III) chain #status predicted <MAT>  
 F;1236-1464/Domain: fibrillar collagen carboxyl-terminal homology <FCC>

Query Match 93.5%; Score 58; DB 2; Length 1464;  
 Best Local Similarity 83.3%; Pred. No. 0.23;  
 Matches 10; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 IAGIGGEKAGGF 12  
 |||||  
 Db 1203 IAGVGGEKSGGF 1214

RESULT 4  
 S14998  
 heat shock protein 18 (clone c9) - maize  
 C;Species: Zea mays (maize)  
 C;Date: 13-Jan-1995 #sequence revision 13-Jan-1995 #text\_change 13-Aug-1999  
 A;Accession: S14998  
 R;Goping, I.S.; Frappier, J.R.H.; Walden, D.B.; Atkinson, B.G.  
 Plant Mol. Biol. 15, 699-711, 1991  
 A;Title: Sequence, identification and characterization of cDNAs encoding two different me  
 A;Reference number: S14997; MUID:91329703; PMID:1714322  
 A;Accession: S14998  
 A;Molecule type: mRNA  
 A;Residues: 1-161 <GOP>  
 A;Cross-references: EMBL:X54076; NID:G22338; PIDN:CAA38013.1; PID:G22339  
 C;Genetics:  
 A;Gene: hsp18-9  
 C;Superfamily: alpha-crystallin  
 C;Keywords: heat shock; stress-induced protein

Query Match 71.0%; Score 44; DB 2; Length 161;  
 Best Local Similarity 80.0%; Pred. No. 5.5;  
 Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 2 AGIGGEKAGG 11  
 |||||  
 Db 27 AGAGGDKAGG 36

RESULT 5  
 F70863  
 probable folicypolyglutamate synthase - Mycobacterium tuberculosis (strain H37RV)  
 C;Species: Mycobacterium tuberculosis  
 C;Date: 17-Jul-1998 #sequence revision 17-Jul-1998 #text\_change 27-Nov-2001

C;Accession: F70863  
 R;Cole, S.T.; Broch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S. Nature 393, 537-544, 1998  
 A;Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
 A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome  
 A;Reference number: A70500; MUID:98295987; PMID:9634230  
 A;Accession: F70863  
 A;Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A;Molecule type: DNA  
 A;Residues: 1-487 <COL>  
 A;Cross-references: GB:AL021246; GB:AL123456; NID:G3261507; PIDN:CAA16024.1; PID:e123757  
 A;Experimental source: strain H37Rv  
 C;Genetics:  
 A;Gene: folC  
 C;Superfamily: folypolyglutamate synthase

Query Match 71.0%; Score 44; DB 2; Length 487;  
 Best Local Similarity 90.0%; Pred. No. 15;  
 Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 IAGIGGEKAG 10  
 |||||  
 Db 209 IAGIGGEKAG 218

RESULT 6  
 AE2981  
 arylesterase [imported] - Agrobacterium tumefaciens (strain C58, Dupont)  
 C;Species: Agrobacterium tumefaciens  
 C;Date: 11-Jan-2002 #sequence\_revision 11-Jan-2002 #text\_change 18-Nov-2002  
 C;Accession: AE2981  
 R;Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I.; erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutayavin, T.; Levy, R.; Li, M.; McClellan; Karp, P.; Romero, P.; Zhang, S. Science 294, 2317-2323, 2001  
 A;Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm, ster, E.W.  
 A;Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.  
 A;Reference number: AB2577; MUID:21608550; PMID:11743193  
 A;Accession: AE2981  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-209 <KUR>  
 A;Cross-references: GB:AE008689; PIDN:AAL44267.1; PID:gl7741852; GSPDB:GN00187  
 A;Experimental source: strain C58 (Dupont)  
 C;Genetics:  
 A;Gene: ada  
 A;Map position: linear chromosome

Query Match 67.7%; Score 42; DB 2; Length 209;  
 Best Local Similarity 72.7%; Pred. No. 15;  
 Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 IAGIGGEKAGG 11  
 :|||  
 Db 137 VAGGGEPPAGG 147

RESULT 7  
 C98302  
 arylesterase (AF044683) [imported] - Agrobacterium tumefaciens (strain C58, Cereon)  
 C;Species: Agrobacterium tumefaciens  
 C;Date: 22-Oct-2001 #sequence\_revision 22-Oct-2001 #text\_change 18-Nov-2002  
 C;Accession: C98302  
 R;Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman, A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.; Science 294, 2323-2328, 2001  
 A;Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tumefaciens C58  
 A;Reference number: A97359; MUID:21608551; PMID:11743194  
 A;Accession: C98302  
 A;Status: preliminary

A;Molecule type: DNA  
 A;Residues: 1-209 <KUR>  
 A;Cross-references: GB:AB007870; PIDN:AAK89941.1; PID:gl5159896; GSPDB:GN00170  
 C;Genetics:  
 A;Gene: AGR L.2749  
 A;Map position: linear chromosome

Query Match 67.7%; Score 42; DB 2; Length 209;  
 Best Local Similarity 72.7%; Pred. No. 15;  
 Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 IAGIGGEKAGG 11  
 :|||  
 Db 137 VAGGGEPPAGG 147

RESULT 8  
 G72777  
 Probable S2P metalloproteinase APE0209 - Aeropyrum pernix (strain K1)  
 C;Species: Aeropyrum pernix  
 C;Date: 20-Aug-1999 #sequence\_revision 20-Aug-1999 #text\_change 20-Aug-1999  
 C;Accession: G72777  
 R;Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Takahawa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; K. DNA Res. 6, 83-101, 1999  
 A;Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyrum pernix strain K1  
 A;Reference number: A72450; MUID:99310339; PMID:10382966  
 A;Accession: G72777  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-383 <KAW>  
 A;Cross-references: DDBJ:AP000058; NID:G5103388; PIDN:BAA79121.1; PID:dl042897; PID:gs101  
 A;Experimental source: strain K1  
 C;Genetics:  
 A;Gene: APE0209

Query Match 66.1%; Score 41; DB 2; Length 383;  
 Best Local Similarity 80.0%; Pred. No. 38;  
 Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 3 GIGGERAGGF 12  
 |||||  
 Db 97 GEGGEAAGGF 106

RESULT 9  
 F69732  
 PBSX prophage ORF xkdo - Bacillus subtilis  
 C;Species: Bacillus subtilis  
 C;Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 15-Oct-1999  
 C;Accession: F69732  
 R;Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Bertero, C.; Bron, S.; Brouillet, S.; Bruschini, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Choi, A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabbet, C.; Ferrari, E. Nature 390, 249-256, 1997  
 A;Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gallucci, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.; Koetter, P.; Koningstein, G.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois, A.; Maueel, Y. M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle, Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon, A.; Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Serot, A.; Winters, P.; Wipat, A.; Yamamoto, K.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K. A;Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.  
 A;Reference number: A69580; MUID:98044033; PMID:9384377  
 A;Accession: F69732  
 A;Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A;Molecule type: DNA  
 A;Residues: 1-1332 <KUN>  
 A;Cross-references: GB:Z99110; GB:AL009126; NID:G2633472; PIDN:CAB13125.1; PID:e1183288;  
 A;Experimental source: strain 168

C;Genetics:  
A;Gene: xkdo

Query Match 66.1%; Score 41; DB 2; Length 1332;  
Best Local Similarity 70.0%; Pred. No. 1.2e+02;  
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 IAGIGGEKAG 10  
: |||||:  
Db 469 VGGIGGEAG 478

## RESULT 10

A37860  
C;Species: Cyprinus carpio (common carp)  
C;Date: 31-May-1991 #sequence\_revision 22-Oct-1999 #text\_change 22-Oct-1999  
C;Accession: A37860  
R;Grabner, M.; Friedrich, K.; Knaus, H.G.; Striesnig, J.; Scheffauer, F.; Staudinger, R.  
Proc. Natl. Acad. Sci. U.S.A. 88, 727-731, 1991  
A;Title: Calcium channels from Cyprinus carpio skeletal muscle.  
A;Reference number: A37860; MUID:91126068; PMID:1846962  
A;Accession: A37860

A;Status: not compared with conceptual translation

A;Molecule type: mRNA

A;Residues: 1-1852 <GRA>

A;Cross-references: GB:M62554; GB:M37203; NID:G213049; PIDN:AAA49205.1; PID:G213050

C;Superfamily: voltage-dependent calcium channel protein alpha-1 chain

C;Keywords: glycoprotein; phosphoprotein; skeletal muscle; transmembrane protein

F;74-90/Domain: transmembrane #status predicted <TR01>  
F;108-131/Domain: transmembrane #status predicted <TR02>  
F;140-158/Domain: transmembrane #status predicted <TR03>  
F;212-234/Domain: transmembrane #status predicted <TR05>  
F;328-350/Domain: transmembrane #status predicted <TR06>  
F;448-466/Domain: transmembrane #status predicted <TR07>  
F;484-501/Domain: transmembrane #status predicted <TR08>  
F;514-530/Domain: transmembrane #status predicted <TR09>  
F;577-596/Domain: transmembrane #status predicted <TR11>  
F;650-676/Domain: transmembrane #status predicted <TR12>  
F;817-834/Domain: transmembrane #status predicted <TR13>  
F;853-870/Domain: transmembrane #status predicted <TR14>  
F;884-901/Domain: transmembrane #status predicted <TR15>  
F;947-966/Domain: transmembrane #status predicted <TR17>  
F;1057-1084/Domain: transmembrane #status predicted <TR18>  
F;1135-1153/Domain: transmembrane #status predicted <TR19>  
F;1169-1188/Domain: transmembrane #status predicted <TR20>  
F;1197-1215/Domain: transmembrane #status predicted <TR21>  
F;1291-1310/Domain: transmembrane #status predicted <TR23>  
F;1377-1402/Domain: transmembrane #status predicted <TR24>  
F;199,102,274,470,813,1157,1269,1485,1703,1713,1745,1760,1848/Binding site: carbohydrate  
F;407/Binding site: phosphate (Thr) (covalent) (by CAMP-dependent kinase) #status predicted  
F;1471,1523,1738/Binding site: phosphate (Ser) (covalent) (by CAMP-dependent kinase) #st

Query Match 66.1%; Score 41; DB 2; Length 1852;  
Best Local Similarity 66.7%; Pred. No. 1.6e+02;  
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 1 IAGIGGEKAGG 12  
: |||||:  
Db 164 IAGVTEKGGF 175

## RESULT 11

H69074  
formylmethanofuran dehydrogenase (EC 1.2.99.5) (tungsten) chain C - Methanobacterium the  
C;Species: Methanobacterium thermoautotrophicum  
C;Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 24-Sep-1999  
C;Accession: H69074  
R;Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.; Dubois, J.; Aldredge, T.;  
Qiu, D.; Spadafora, R.; Vicaire, R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiwan, N.  
ki, S.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.  
J. Bacteriol. 179, 7135-7155, 1997  
A;Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta H: funcn

A;Reference number: A69000; MUID:98037514; PMID:9371463

A;Accession: H69074

A;Status: preliminary; nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-270 <MTH>

A;Cross-references: GB:AE000916; GB:AE000666; NID:G2622674; PIDN:AAB86032.1; PID:G262268

A;Experimental source: strain Delta H

C;Genetics:

A;Gene: MTH1558

C;Superfamily: formylmethanofuran dehydrogenase chain C

C;Keywords: iron-sulfur protein; metalloprotein; oxidoreductase; tungsten

Query Match 64.5%; Score 40; DB 2; Length 270;

Best Local Similarity 63.6%; Pred. No. 40;

Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 IAGIGGEKAGG 11  
: |||||:  
Db 197 VARVGGEWAGG 207

## RESULT 12

S57457  
formylmethanofuran dehydrogenase (EC 1.2.99.5) (tungsten) chain C - Methanobacterium the  
C;Species: Methanobacterium thermoautotrophicum  
C;Date: 10-Oct-1995 #sequence\_revision 17-Nov-1995 #text\_change 20-Jun-2000  
C;Accession: S63546; S57457  
R;Hochheimer, A.; Schmitz, R.A.; Thauer, R.K.; Hedderich, R.  
Eur. J. Biochem. 234, 910-920, 1995

A;Title: The tungsten formylmethanofuran dehydrogenase from Methanobacterium thermoautotot

A;Reference number: S63519; MUID:96163477; PMID:8575452

A;Accession: S63546

A;Molecule type: DNA

A;Residues: 1-270 <HOC>

A;Cross-references: EMBL:X87970; NID:G1890205; PIDN:CAA61214.1; PID:G1890211

A;Experimental source: strain Marburg, DSM 2133

C;Genetics:

A;Gene: fwdC

C;Complex: heterotetramer of A (see PIR:S57456), B (see PIR:S57458), C, and D (see PIR:S5

C;Superfamily: formylmethanofuran dehydrogenase chain C

C;Keywords: heterotetramer; iron-sulfur protein; metalloprotein; oxidoreductase; tungster

Query Match 64.5%; Score 40; DB 2; Length 270;

Best Local Similarity 63.6%; Pred. No. 40;

Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 IAGIGGEKAGG 11  
: |||||:  
Db 197 VARVGGEWAGG 207

## RESULT 13

E95286  
hypothetical protein Sma0367 [imported] - Sinorhizobium meliloti (strain 1021) magaplasmi  
C;Species: Sinorhizobium meliloti  
C;Date: 24-Aug-2001 #sequence\_revision 24-Aug-2001 #text\_change 30-Sep-2001  
C;Accession: E95286  
R;Barnett, M.J.; Fisher, R.F.; Jones, T.; Komp, C.; Abola, A.P.; Barloy-Hubler, F.; Bowse  
; Kalman, S.; Keating, D.H.; Palm, C.; Peck, M.C.; Surzycki, R.; Wells, D.H.; Yeh, K.C.;  
Proc. Natl. Acad. Sci. U.S.A. 98, 9883-9888, 2001

A;Title: Nucleotide sequence and predicted functions of the entire Sinorhizobium meliloti

A;Reference number: A95262; MUID:21396509; PMID:11481432

A;Accession: E95286

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-298 <KUR>

A;Cross-references: GB:AE006469; PIDN:AAK64855.1; PID:G14523269; GSPDB:GN00165

A;Experimental source: strain 1021, megaplasmid pSymA

R;Galibert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubler,  
pela, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.;  
L.; Hyman, R.W.; Jones, T.

Science 293, 668-672, 2001

A;Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelaure,

hebault, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K.  
 A;Title: The composite genome of the legume symbiont *Sinorhizobium meliloti*.  
 A;Reference number: A96039; MUID:21368234; PMID:11474104  
 A;Contents: annotation  
 C;Genetics:  
 A;Gene: Sma0367  
 A;Genome: plasmid

Query Match 64.5%; Score 40; DB 2; Length 298;  
 Best Local Similarity 88.9%; Pred. No. 43;  
 Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 AGIGGEKAG 10  
 |||||  
 Db 103 AGIGGAKAG 111

## RESULT 14

C64654  
 conserved hypothetical secreted protein HP1075 - *Helicobacter pylori* (strain 26695)  
 C;Species: *Helicobacter pylori*  
 C;Date: 09-Aug-1997 #sequence\_revision 09-Aug-1997 #text\_change 18-Aug-2000  
 C;Accession: C64654  
 R;Tomb, J.P.; White, O.; Karlavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.D.; Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khalak, H.G.; Glodek, A.; McKenna, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Watthey, L. Nature 388, 539-547, 1997  
 A;Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser, C.  
 A;Title: The complete genome sequence of the gastric pathogen *Helicobacter pylori*.  
 A;Reference number: A64520; MUID:97394467; PMID:9252185  
 A;Accession: C64654  
 A;Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A;Molecule type: DNA  
 A;Residues: 1-438 <TOM>  
 A;Cross-references: GB:AE000614; GB:AE000511; NID:g2314216; PIDN:AAD08118.1; PID:g231422  
 C;Superfamily: *Helicobacter pylori* hypothetical protein jhp0350

Query Match 64.5%; Score 40; DB 2; Length 438;  
 Best Local Similarity 58.3%; Pred. No. 62;  
 Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Qy 1 IAGIGGEKAGGF 12  
 :||| |:: |||  
 Db 42 LAGIQGDEPGGF 53

## RESULT 15

H71943  
 hypothetical protein jhp0350 - *Helicobacter pylori* (strain J99)  
 C;Species: *Helicobacter pylori*  
 A;Variety: strain J99  
 C;Date: 12-Feb-1999 #sequence\_revision 12-Feb-1999 #text\_change 18-Aug-2000  
 C;Accession: H71943  
 R;Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D.R.; Ives, C.; Gibson, R.; Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F.; Nature 397, 176-180, 1999  
 A;Title: Genomic sequence comparison of two unrelated isolates of the human gastric pathogen *Helicobacter pylori*.  
 A;Reference number: A71800; MUID:99120557; PMID:9923682  
 A;Accession: H71943  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-438 <ARN>  
 A;Cross-references: GB:AE001470; GB:AE001439; NID:g4154869; PIDN:AAD05927.1; PID:g415487  
 A;Experimental source: strain J99  
 C;Genetics:  
 A;Gene: jhp0350  
 C;Superfamily: *Helicobacter pylori* hypothetical protein jhp0350

Query Match 64.5%; Score 40; DB 2; Length 438;  
 Best Local Similarity 58.3%; Pred. No. 62;  
 Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Qy 1 IAGIGGEKAGGF 12

Db :||| |:: |||  
 42 LAGIQGDEPGGF 53

Search completed: September 18, 2004, 04:30:54  
 Job time : 15.5882 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 18, 2004, 03:48:54 ; Search time 8.94118 Seconds  
(without alignments)  
69.884 Million cell updates/sec

Title: US-10-615-959-43  
Perfect score: 62  
Sequence: 1 IAGIGGEKAGGF 12

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_42.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	62	100.0	1466	1 CAL3 HUMAN	P02461 homo sapien
2	58	93.5	636	1 CAL3 RAT	P13941 ratus norv
3	58	93.5	1464	1 CAL3 MOUSE	P08121 mus musculu
4	44	71.0	161	1 HS21 MAIZE	P24631 zea mays (m
5	41	66.1	1332	1 XKDO BACSU	P54334 bacillus su
6	41	66.1	1852	1 CCAS CYPCA	P22316 cyprinus ca
7	40	64.5	270	1 FWDC METTH	O27600 methanobact
8	40	64.5	270	1 FWDC METTM	Q59579 methanobact
9	40	64.5	383	1 DNAT PORGI	Q9XCA6 porphyronon
10	39	62.9	175	1 YOCAL CAEEL	Q23280 caenorhabdi
11	39	62.9	354	1 RFL LEPIN	Q819S6 leptospira
12	39	62.9	378	1 CSP PLACL	P08675 plasmodium
13	39	62.9	399	1 HXAA MOUSE	P31310 mus musculu
14	39	62.9	1679	1 FUR2 DROME	P30432 drosophila
15	38	61.3	267	1 MIRC METAC	Q8TU01 methanosarc
16	38	61.3	400	1 LEUK HUMAN	P16150 homo sapien
17	38	61.3	423	1 TBX2 CAEEL	Q19691 caenorhabdi
18	38	61.3	495	1 AMPA PSEAE	O68822 pseudomonas
19	38	61.3	592	1 RB56 HUMAN	Q92804 homo sapien
20	37	59.7	62	1 LAFX IACJO	Q48509 lactobacill
21	37	59.7	198	1 FLIN RHIME	O54245 rhizobium m
22	37	59.7	232	1 SSB COREF	Q8FIP9 corynebacte
23	37	59.7	269	1 POLG PWVMI	P32574 passionfrui
24	37	59.7	371	1 TRMU VERPE	Q8ZFQ5 versinia pe
25	37	59.7	384	1 GRP1 PETHY	P09789 petunia hyb
26	37	59.7	388	1 NAH3 METJA	O58671 methanococc
27	37	59.7	422	1 VDR XENLA	O13124 xenopus lae
28	37	59.7	462	1 TRPE THETH	P05378 thermus lae
29	37	59.7	468	1 IL9R MOUSE	Q01114 mus musculu
30	37	59.7	471	1 K1CN HUMAN	P02533 homo sapien
31	37	59.7	486	1 XYLQ PSEPU	P23105 pseudomonas
32	37	59.7	499	1 DLDH YEAST	P09624 saccharomyc
33	37	59.7	562	1 NPRM_BACME	Q00891 bacillus me

#### RESULT 1

ID	CAL3 HUMAN	STANDARD;	PRT;	1466 AA.
AC	P02461; Q15112;			
DT	21-JUL-1986 (Rel. 01, Created)			
DT	01-JAN-1990 (Rel. 13, Last sequence update)			
DT	10-OCT-2003 (Rel. 42, Last annotation update)			
DE	Collagen alpha 1(III) chain precursor.			
GN	COL3A1.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Skin fibroblast;			
RX	MEDLINE=89350838; PubMed=2764886;			
RA	Ala-Kokko L., Kontusaari S., Baldwin C.T., Kuivaniemi H.,			
RA	Prockop D.J.;			
RT	"Structure of cDNA clones coding for the entire prepro alpha 1 (III) chain of human type III procollagen. Differences in protein structure from type I procollagen and conservation of codon preferences.";			
RT	Biochem. J. 260:509-516(1989).			
RN	[2]			
RP	SEQUENCE OF 149-1225 FROM N.A.			
RX	MEDLINE=89386015; PubMed=2780304;			
RA	Janecko R.A., Ramirez F.;			
RT	"Nucleotide and amino acid sequences of the entire human alpha 1 (III) collagen.";			
RT	Nucleic Acids Res. 17:6742-6742(1989).			
RN	[3]			
RP	SEQUENCE OF 168-398.			
RX	MEDLINE=77134724; PubMed=557335;			
RA	Seyer J.M., Kang A.H.;			
RT	"Covalent structure of collagen: amino acid sequence of cyanogen bromide peptides from the amino-terminal segment of type III collagen of human liver.";			
RT	Biochemistry 16:1158-1164(1977).			
RN	[4]			
RP	REVISIONS.			
RA	Seyer J.M.;			
RL	Submitted (DEC-1977) to the PIR data bank.			
RN	[5]			
RP	SEQUENCE OF 399-727.			
RX	MEDLINE=79000343; PubMed=687591;			
RA	Seyer J.M., Kang A.H.;			
RT	"Covalent structure of collagen: amino acid sequence of five consecutive CNBR peptides from type III collagen of human liver.";			
RT	Biochemistry 17:3404-3411(1978).			
RN	[6]			
RP	SEQUENCE OF 728-964.			
RX	MEDLINE=80198282; PubMed=6246925;			
RA	Seyer J.M., Mainardi C., Kang A.H.;			
RT	"Covalent structure of collagen: amino acid sequence of alpha 1 (III)-CB5 from type III collagen of human liver.";			
RT	Biochemistry 19:1583-1589(1980).			

#### ALIGNMENTS

ID	CAL3 HUMAN	STANDARD;	PRT;	1466 AA.
AC	P02461; Q15112;			
DT	21-JUL-1986 (Rel. 01, Created)			
DT	01-JAN-1990 (Rel. 13, Last sequence update)			
DT	10-OCT-2003 (Rel. 42, Last annotation update)			
DE	Collagen alpha 1(III) chain precursor.			
GN	COL3A1.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Skin fibroblast;			
RX	MEDLINE=89350838; PubMed=2764886;			
RA	Ala-Kokko L., Kontusaari S., Baldwin C.T., Kuivaniemi H.,			
RA	Prockop D.J.;			
RT	"Structure of cDNA clones coding for the entire prepro alpha 1 (III) chain of human type III procollagen. Differences in protein structure from type I procollagen and conservation of codon preferences.";			
RT	Biochem. J. 260:509-516(1989).			
RN	[2]			
RP	SEQUENCE OF 149-1225 FROM N.A.			
RX	MEDLINE=89386015; PubMed=2780304;			
RA	Janecko R.A., Ramirez F.;			
RT	"Nucleotide and amino acid sequences of the entire human alpha 1 (III) collagen.";			
RT	Nucleic Acids Res. 17:6742-6742(1989).			
RN	[3]			
RP	SEQUENCE OF 168-398.			
RX	MEDLINE=77134724; PubMed=557335;			
RA	Seyer J.M., Kang A.H.;			
RT	"Covalent structure of collagen: amino acid sequence of cyanogen bromide peptides from the amino-terminal segment of type III collagen of human liver.";			
RT	Biochemistry 16:1158-1164(1977).			
RN	[4]			
RP	REVISIONS.			
RA	Seyer J.M.;			
RL	Submitted (DEC-1977) to the PIR data bank.			
RN	[5]			
RP	SEQUENCE OF 399-727.			
RX	MEDLINE=79000343; PubMed=687591;			
RA	Seyer J.M., Kang A.H.;			
RT	"Covalent structure of collagen: amino acid sequence of five consecutive CNBR peptides from type III collagen of human liver.";			
RT	Biochemistry 17:3404-3411(1978).			
RN	[6]			
RP	SEQUENCE OF 728-964.			
RX	MEDLINE=80198282; PubMed=6246925;			
RA	Seyer J.M., Mainardi C., Kang A.H.;			
RT	"Covalent structure of collagen: amino acid sequence of alpha 1 (III)-CB5 from type III collagen of human liver.";			
RT	Biochemistry 19:1583-1589(1980).			

RA [7] SEQUENCE OF 950-1466 FROM N.A.  
 RP MEDLINE=88189827; PubMed=3357782;  
 RX Mankoo B.S., Dalglish R.;  
 RA "Human pro alpha 1(III) collagen: cDNA sequence for the 3' end.";  
 RT Nucleic Acids Res. 16:2337-2337(1988).  
 RL [8]  
 RN  
 RP REVISION TO 1184.  
 RX MEDLINE=89098346; PubMed=3211760;  
 RA Molyneux K., Dalglish R.;  
 RT "Human type III collagen 'variant' is a cDNA cloning artefact.";  
 RL Nucleic Acids Res. 16:11833-11833(1988).  
 RN [9]  
 RP SEQUENCE OF 1065-1466 FROM N.A.  
 RX MEDLINE=85087944; PubMed=6096827;  
 RA Loidi H.R., Brinker J.M., May M., Pihlajaniemi T., Morrow S.,  
 RT Rosenbloom J., Myers J.C.;  
 RL "Molecular cloning and carboxyl-propeptide analysis of human type III  
 RT procollagen.";  
 RL Nucleic Acids Res. 12:9383-9394(1984).  
 RN [10]  
 RP SEQUENCE OF 965-1200.  
 RX MEDLINE=81208139; PubMed=7016180;  
 RA Seyer J.M., Kang A.H.;  
 RT "Covalent structure of collagen: amino acid sequence of alpha  
 RT 1(III)-CB9 from type III collagen of human liver.";  
 RL Biochemistry 20:2621-2627(1981).  
 RN [11]  
 RP SEQUENCE OF 1176-1466 FROM N.A.  
 RX MEDLINE=85157600; PubMed=2579949;  
 RA Chu M.-L., Weil D., de Wet W.J., Bernard M.P., Sippola M., Ramirez F.;  
 RT "Isolation of cDNA and genomic clones encoding human pro-alpha 1  
 RT (III) collagen. Partial characterization of the 3' end region of the  
 RT gene.";  
 RL J. Biol. Chem. 260:4357-4363(1985).  
 RN [12]  
 RP SEQUENCE OF 1161-1200 FROM N.A.  
 RX MEDLINE=86187804; PubMed=3754462;  
 RA Miskulin M., Dalglish R., Kluge-Beckerman B., Rennard S.I.,  
 RT Tolstoshev P., Brantly M., Crystal R.G.;  
 RT "Human type III collagen gene expression is coordinately modulated  
 RT with the type I collagen genes during fibroblast growth.";  
 RL Biochemistry 25:1408-1413(1986).  
 RN [13]  
 RP SEQUENCE OF 1-170 FROM N.A.  
 RX TISSUE=Placenta;  
 RA Tomán D., Ricca G., de Crombrughe B.;  
 RT "Nucleotide sequence of a cDNA coding for the amino-terminal region  
 RT of human prepro alpha 1(III) collagen.";  
 RL Nucleic Acids Res. 16:7201-7201(1988).  
 RN [14]  
 RP SEQUENCE OF 1-176 FROM N.A.  
 RX MEDLINE=89378752; PubMed=27777083;  
 RA Benson-Chanda V., Su M.W., Weil D., Chu M.-L., Ramirez F.;  
 RT "Cloning and analysis of the 5' portion of the human type-III  
 RT procollagen gene (COL3A1).";  
 RL Gene 78:255-265(1989).  
 RN [15]  
 RP REVIEW ON VARIANTS.  
 RX MEDLINE=97255959; PubMed=9101290;  
 RA Kuivaniemi H., Tromp G., Prockop D.J.;  
 RT "Mutations in fibrillar collagens (types I, II, III, and XI), fibril-  
 RT associated collagen (type IX), and network-forming collagen (type X)  
 RT cause a spectrum of diseases of bone, cartilage, and blood vessels.";  
 RL Hum. Mutat. 9:300-315(1997).  
 RN [16]  
 RP VARIANT AORTIC ANEURYSM ARG-303, AND VARIANT THR-668.  
 RX MEDLINE=93293988; PubMed=8514866;  
 RA Tromp G., Wu Y., Prockop D.J., Madhathari S.L., Kleinert C.,  
 RA Earley J.J., Zhuang J., Noerregaard O., Darling R.C., Abbott W.M.,  
 RA Cole C.W., Jaakkola P., Ryyanen M., Pearce W.H., Yao J.S.T.,  
 RA Majamaa K., Smullens S.N., Gatalica Z., Ferrell R.E., Jimenez S.A.,  
 RA Jackson C.E., Michels V.V., Kaye M., Kuivaniemi H.;  
 RT "Sequencing of cDNA from 50 unrelated patients reveals that mutations  
 RT in the triple-helical domain of type III procollagen are an  
 RT infrequent cause of aortic aneurysms.";  
 RL J. Clin. Invest. 91:2539-2545(1993).  
 RN [17]  
 RP VARIANT THR-698.  
 RX MEDLINE=91045136; PubMed=2235526;  
 RA Zafarullah K., Kleinert C., Tromp G., Kuivaniemi H., Kontusaari S.,  
 RA Wu Y., Ganguly A., Prockop D.J.;  
 RT "G to A polymorphism in exon 31 of the COL3A1 gene.";  
 RL Nucleic Acids Res. 18:6180-6180(1990).  
 RN [18]  
 RP VARIANT AORTIC ANEURYSM ARG-786.  
 RX MEDLINE=91056145; PubMed=2243125;  
 RA Kontusaari S., Tromp G., Kuivaniemi H., Romanic A.M., Prockop D.J.;  
 RT "A mutation in the gene for type III procollagen (COL3A1) in a family  
 RT with aortic aneurysms.";  
 RL J. Clin. Invest. 86:1465-1473(1990).  
 RN [19]  
 RP VARIANT EDS-IV ARG-828.  
 RX MEDLINE=94016385; PubMed=8411057;  
 RA Richards A.J., Narcisi P., Lloyd J.C., Ferguson C., Pope F.M.;  
 RT "The substitution of glycine 661 by arginine in type III collagen  
 RT produces mutant molecules with different thermal stabilities and  
 RT causes Ehlers-Danlos syndrome type IV.";  
 RL J. Med. Genet. 30:690-693(1993).  
 RN [20]  
 RP VARIANT EDS-IV SER-957.  
 RX MEDLINE=89109135; PubMed=2492273;  
 RA Tromp G., Kuivaniemi H., Shikata H., Prockop D.J.;  
 RT "A single base mutation that substitutes serine for glycine 790 of  
 RT the alpha 1 (III) chain of type III procollagen exposes an arginine  
 RT and causes Ehlers-Danlos syndrome IV.";  
 RL J. Biol. Chem. 264:1349-1352(1989).  
 RN [21]  
 RP VARIANT EDS-IV VAL-960.  
 RX MEDLINE=95268429; PubMed=7749417;  
 RA Tromp G., de Paeppe A., Nuytinck L., Madhathari S.L., Kuivaniemi H.;  
 RT "Substitution of valine for glycine 793 in type III procollagen in  
 RT Ehlers-Danlos syndrome type IV.";  
 RL Hum. Mutat. 5:179-181(1995).  
 RN [22]  
 RP VARIANT EDS-IV GLU-1014.  
 RX MEDLINE=92316511; PubMed=1352273;  
 RA Richards A.J., Ward P.N., Narcisi P., Nicholls A.C., Lloyd J.C.,  
 RA Pope F.M.;  
 RT "A single base mutation in the gene for type III collagen (COL3A1)  
 RT converts glycine 847 to glutamic acid in a family with Ehlers-Danlos  
 RT syndrome type IV. An unaffected family member is mosaic for the  
 RT mutation.";  
 RL Hum. Genet. 89:414-418(1992).  
 RN [23]  
 RP VARIANT EDS-IV ASP-1050.  
 RX MEDLINE=90037070; PubMed=2808425;  
 RA Tromp G., Kuivaniemi H., Stolle C.A., Pope F.M., Prockop D.J.;  
 RT "Single base mutation in the type III procollagen gene that converts  
 RT the codon for glycine 883 to aspartate in a mild variant of  
 RT Ehlers-Danlos syndrome IV.";  
 RL J. Biol. Chem. 264:19313-19317(1989).  
 RN [24]  
 RP VARIANT EDS-IV VAL-1077.  
 RX MEDLINE=91374480; PubMed=1895316;  
 RA Richards A.J., Lloyd J.C., Ward P.N., de Paeppe A., Narcisi P.,  
 RA Pope F.M.;  
 RT "Characterisation of a glycine to valine substitution at amino acid  
 RT position 910 of the triple helical region of type III collagen in a  
 RT patient with Ehlers-Danlos syndrome type IV.";  
 RL J. Med. Genet. 28:458-463(1991).  
 RN [25]  
 RP VARIANT EDS-IV GLU-1173.  
 RX MEDLINE=93022543; PubMed=1357232;  
 RA Johnson P.H., Richards A.J., Pope F.M., Hopkinson D.A.;

Query Match 100.0%; Score 62; DB 1; Length 1466;  
 Best Local Similarity 100.0%; Pred. No. 0.032;  
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IAGIGGKAGGF 12  
 DB 1205 IAGIGGKAGGF 1216  
 |||||

RESULT 2  
 CA13\_RAT STANDARD; PRT; 636 AA.  
 ID P13941; OT0604;  
 AC 01-JAN-1990 (Rel. 13, Created)  
 DT 01-JUN-1994 (Rel. 23, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Collagen alpha 1(III) chain (Fragment).  
 GN COL3A1.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=94114571; PubMed=8286415;  
 RA Glumoff V., Maekele J.K., Vuorio E.;  
 RT "Cloning of cDNA for rat pro alpha 1(III) collagen mRNA. Different  
 RT expression patterns of type I and type III collagen and fibronectin  
 RT genes in experimental granulation tissue.";  
 RL Biochim. Biophys. Acta 1217:41-48(1994).  
 RN [2]  
 RP SEQUENCE OF 73-636 FROM N.A.  
 RC STRAIN=Sprague-Dawley; TISSUE=Fibroblast;  
 RA Wurtz T., Ellerstrom C., Lundmark C., Christersson C.;  
 RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE OF 308-482 FROM N.A.  
 RX MEDLINE=88296083; PubMed=2456904;  
 RA Frankel F.R., Hsu C.-Y.J., Meyers J.C., Lin E., Lytle C.R.,  
 RA Kohn M., Mohn K.;  
 RT "Regulation of alpha 2(I), alpha 1(III), and alpha 2(V) collagen  
 RT mRNAs by estradiol in the immature rat uterus.";  
 RL DNA 7:347-354(1988).  
 CC -!- FUNCTION: Collagen type III occurs in most soft connective tissues  
 CC along with type I collagen.  
 CC -!- SUBUNIT: Trimers of identical alpha 1(III) chains. The chains are  
 CC linked to each other by interchain disulfide bonds. Trimers are  
 CC also cross-linked via hydroxyllysines.  
 CC -!- PTM: Prolines at the third position of the tripeptide repeating  
 CC unit (G-X-Y) are hydroxylated in some or all of the chains.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; X70369; CAA49832.1; -;  
 DR EMBL; AJ005395; CAA06510.1; -;  
 DR EMBL; M21354; AAA40942.1; -;  
 DR FIR; S41067; S41067.  
 DR InterPro; IPR008161; C1g helix.  
 DR InterPro; IPR008160; Collagen.  
 DR InterPro; IPR000885; Fib collagen\_C.  
 DR InterPro; IPR002181; Fibrinogen\_C.  
 DR InterPro; IPR001007; VWF C.  
 DR Pfam; PF01410; COLFI; 1.  
 DR ProDom; PD000007; C1g helix; 1.  
 DR ProDom; PD002078; Fib\_collagen\_C; 1.

DR SMART; SM00038; COLFI; 1.  
 DR PROSITE; PS01208; WVFC\_1; PARTIAL.  
 KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;  
 KW Collagen; Glycoprotein.  
 FT NON\_TER 1  
 FT CHAIN <1 375  
 FT PROPEP 376 636  
 FT DOMAIN <1 368  
 FT DISULFID 369 368  
 FT DISULFID 369 368  
 FT CONFLICT 340 340  
 FT CONFLICT 429 429  
 SQ SEQUENCE 636 AA; 62332 MW; 61A48159F01D01EE CRC64;  
 Query Match 93.5%; Score 58; DB 1; Length 636;  
 Best Local Similarity 83.3%; Pred. No. 0.066;  
 Matches 10; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 IAGIGGKAGGF 12  
 DB 375 IAGVGGEKSGGF 386  
 |||||

RESULT 3  
 CA13\_MOUSE STANDARD; PRT; 1464 AA.  
 ID P08121; Q61429; Q9CRN7;  
 AC 01-AUG-1988 (Rel. 08, Created)  
 DT 15-JUL-1999 (Rel. 38, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE Collagen alpha 1(III) chain precursor.  
 GN COL3A1.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6 X DBA; TISSUE=Embryo;  
 RX MEDLINE=95011609; PubMed=7926795;  
 RA Tonan D., de Crombrughe B.;  
 RT "The mouse type-III procollagen-encoding gene: genomic cloning and  
 RT complete DNA sequence.";  
 RL Gene 147:161-168(1994).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6; TISSUE=Brain;  
 RX MEDLINE=22388257; PubMed=12477932;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh P.,  
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Udwin T.B., Toshiyuki S., Carninci P., Frange C.,  
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Pahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Butlerfield Y.S.N., Krzywinski M.I., Skalska U., Smalish D.E.,  
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length  
 RT human and mouse cDNA sequences.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 RN [3]  
 RP SEQUENCE OF 1-488 FROM N.A.  
 RX MEDLINE=88167858; PubMed=3443309;  
 RA Wood L., Theriault N., Vogeli G.;



RT "Complete nucleotide sequence of the N-terminal domains of the murine  
 RT alpha-1 type-III collagen chain.";  
 RL Gene 61:225-230(1987).  
 RN [4]  
 RP SEQUENCE OF 1-28 FROM N.A.  
 RX MEDLINE=85131189; PubMed=3972847;  
 RA Liau G., Mudryj M., de Crombrughe B.;  
 RT "Identification of the promoter and first exon of the mouse alpha 1  
 RT (III) collagen gene.";  
 RL J. Biol. Chem. 260:3773-3777 (1985).  
 RN [5]  
 RP SEQUENCE OF 810-1464 FROM N.A.  
 RX STRAIN=C57BL/6J; TISSUE=Embryonic head;  
 RX MEDLINE=21095660; PubMed=11217851;  
 RA Kawakawa T., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
 RA Arakawa K., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,  
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,  
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,  
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
 RA Fleischmann W., Gaasterland T., Giesi C., King B., Kochiwa H.,  
 RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,  
 RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,  
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,  
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,  
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,  
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,  
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,  
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,  
 RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,  
 RA Hayashizaki Y.;  
 RT "Functional annotation of a full-length mouse cDNA collection.";  
 RL Nature 409:685-690(2001).  
 RN [6]  
 RP SEQUENCE OF 1442-1464 FROM N.A.  
 RX STRAIN=C57BL/6;  
 RX MEDLINE=91274355; PubMed=2054384;  
 RA Metsieranta M., Toman D., de Crombrughe B., Vuorio E.;  
 RT "Specific hybridization probes for mouse type I, II, III and IX  
 RT collagen mRNAs.";  
 RL Biochim. Biophys. Acta 1089:241-243(1991).  
 CC -!- FUNCTION: Collagen type III occurs in most soft connective tissues  
 CC along with type I collagen.  
 CC -!- SUBUNIT: Trimers of identical alpha 1(III) chains. The chains are  
 CC linked to each other by interchain disulfide bonds. Trimers are  
 CC also cross-linked via hydroxyllysines.  
 CC -!- PTM: Proline residues at the third position of the tripeptide  
 CC repeating unit (G-X-Y) are hydroxylated in some or all of the  
 CC chains.  
 CC -!- PTM: O-linked glycan consists of a Glc-Gal disaccharide bound to  
 CC the oxygen atom of a post-translationally added hydroxyl group (By  
 CC similarity).  
 CC -!- SIMILARITY: Contains 1 WFAC domain.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; X52046; CAA36279.1; -;  
 CC EMBL; BC043089; AAH43089.1; -;  
 CC EMBL; BC058724; AAH58724.1; -;  
 CC EMBL; M18933; AAA37338.1; -;  
 CC EMBL; K03037; -; NOT ANNOTATED\_CDS.  
 CC EMBL; AK019448; BAB31724.1; -;  
 CC EMBL; X57983; CAA41048.1; -;  
 CC EMBL; A27353; A27353.  
 CC EMBL; S59856; S59856.  
 CC MGD; MGI:88453; Col3a1.

DR InterPro; IPR008161; Clq\_helix.  
 DR InterPro; IPR008160; Collagen.  
 DR InterPro; IPR00885; Fib\_collagen\_C.  
 DR InterPro; IPR002181; Fibrinogen\_C.  
 DR InterPro; IPR001007; WF\_C.  
 DR Pfam; PF01410; COLFI; 1.  
 DR Pfam; PF01391; Collagen; 18.  
 DR ProDom; PD000007; Clq\_helix; 1.  
 DR ProDom; PD002078; Fib\_collagen\_C; 1.  
 DR SMART; SMO0038; COLFI; 1.  
 DR SMART; SMO0214; VMC; 1.  
 DR PROSITE; PS01208; WFAC; 1.  
 DR PROSITE; PS0184; WFAC; 2; 1.  
 KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;  
 KW Glycoprotein; Collagen; Signal.  
 FT SIGNAL 1 23 BY SIMILARITY.  
 FT PROPEP 24 154 AMINO-TERMINAL PROPEPTIDE.  
 FT CHAIN 155 1203 COLLAGEN ALPHA 1(III) CHAIN.  
 FT PROPEP 1204 1464 CARBOXYL-TERMINAL PROPEPTIDE.  
 FT DOMAIN 31 90 WFAC.  
 FT DOMAIN 155 169 NONHELICAL REGION (N-TERMINAL).  
 FT DOMAIN 170 1195 TRIPLE-HELICAL REGION.  
 FT DOMAIN 1196 1464 NONHELICAL REGION (C-TERMINAL).  
 FT CARBOHYD 262 262 O-LINKED (GAL. . .) (BY SIMILARITY).  
 FT MOD\_RES 262 262 HYDROXYLATION (BY SIMILARITY).  
 FT MOD\_RES 283 283 HYDROXYLATION (BY SIMILARITY).  
 FT MOD\_RES 859 859 HYDROXYLATION (BY SIMILARITY).  
 FT MOD\_RES 976 976 HYDROXYLATION (BY SIMILARITY).  
 FT MOD\_RES 1093 1093 HYDROXYLATION (BY SIMILARITY).  
 FT MOD\_RES 1105 1105 HYDROXYLATION (BY SIMILARITY).  
 FT DISULFID 1195 1195 INTERCHAIN (BY SIMILARITY).  
 FT DISULFID 1196 1196 INTERCHAIN (BY SIMILARITY).  
 SQ SEQUENCE 1464 AA; 138944 MW; 2104EC27A886090B CRC64;  
 Query Match 93.5%; Score 58; DB 1; Length 1464;  
 Best Local Similarity 83.3%; Pred. No. 0.14;  
 Matches 10; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 IAGIGGEKAGGF 12  
 DB 1203 IAGVGGEKSGGF 1214  
 RESULT 4  
 HS21 MAIZE  
 ID HS21 MAIZE STANDARD; PRT; 161 AA.  
 AC P24631;  
 DT 01-MAR-1992 (Rel. 21, Created)  
 DT 01-MAR-1992 (Rel. 21, Last sequence update)  
 DT 01-JUN-1994 (Rel. 29, Last annotation update)  
 DE 17.5 kDa class II heat shock protein.  
 OS Zea mays (Maize).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC PACCAD clade; Panicoideae; Andropogoneae; Zea.  
 OX NCBI\_TaxID=4577;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Ohio 43; TISSUE=Radicle;  
 RX MEDLINE=91329703; PubMed=1714322;  
 RA Goping I.S., Frappier J.R.H., Waldeen D.B., Atkinson B.G.;  
 RT "Sequence, identification and characterization of cDNAs encoding two  
 RT different members of the 18 kDa heat shock family of Zea mays L.";  
 RL Plant Mol. Biol. 16:699-711(1991).  
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic.  
 CC -!- SIMILARITY: Belongs to the small heat shock protein (HSP20)  
 CC family.  
 CC -!- SIMILARITY: PLANTS HAVE FOUR CLASSES OF HSP20: CLASS I AND II  
 CC WHICH ARE CYTOPLASMIC; CLASS III WHICH IS IN THE CHLOROPLAST  
 CC AND CLASS IV WHICH IS IN THE ENOMEMBRANE. THIS PROTEIN BELONGS  
 CC TO CLASS II.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration



CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; X54076; CAA38013.1; -;  
 DR PIR; S14998; S14998.  
 DR MaizeDB; S1309; -;  
 DR InterPro; IPR002068; HSP20.  
 DR InterPro; IPR008978; HSP20\_chap.  
 DR Pfam; PF00011; HSP20; 1.  
 DR PROSITE; PS01031; HSP20; 1.  
 KW Heat shock; Multigene family.  
 SQ SEQUENCE 161 AA; 17511 MW; 3B60A8DD4396577C CRC64;

Query Match 71.0%; Score 44; DB 1; Length 161;  
 Best Local Similarity 80.0%; Pred. No. 3.3;  
 Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
 QY 2 AGIGGEKAGG 11  
 |||||  
 DB 27 AGAGGDKAGG 36

RESULT 5  
 XKDO\_BACSU STANDARD; PRT; 1332 AA.  
 AC P54334;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Phage-like element PBSX protein xkdo.  
 GN XKDO OR BSU12680.  
 OS Bacillus subtilis.  
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
 OX NCBI\_TaxID=1423;  
 (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=168;  
 RA Krogh S., O'Reilly M., Nolan N., Devine K.M.;  
 RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.  
 (2)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=168;  
 RA Kunst P., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,  
 RA Azavedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,  
 RA Borries R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,  
 RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,  
 RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,  
 RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,  
 RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,  
 RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,  
 RA Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,  
 RA Giuseppe G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,  
 RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,  
 RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,  
 RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,  
 RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,  
 RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,  
 RA Medina N., Millaud R.P., Mizuno M., Moestl D., Nakai S., Noback M.,  
 RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudea B., Park S.H.,  
 RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,  
 RA Presacco E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,  
 RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadate Y.,  
 RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,  
 RA Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,  
 RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takenaru K.,  
 RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni K.,  
 RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,  
 RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzenecker T.,

RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,  
 RA Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;  
 RT "The complete genome sequence of the Gram-positive bacterium *Bacillus*  
 RT *subtilis*,"  
 RL Nature 390:249-256(1997).  
 CC -!- SIMILARITY: STRONG, TO B.SUBTILIS YOBO.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; Z70177; CAA94037.1; -;  
 DR EMBL; Z99110; CAB13125.1; -;  
 DR PIR; P69732; F69732.  
 DR Subtilist; BGL1549; xkdo.  
 DR InterPro; IPR008258; SLT\_dom.  
 DR Pfam; PF01464; SLT; 1.  
 KW Complete proteome.  
 SQ SEQUENCE 1332 AA; 145149 MW; A96C9C9F9E31DF01 CRC64;

Query Match 66.1%; Score 41; DB 1; Length 1332;  
 Best Local Similarity 70.0%; Pred. No. 66;  
 Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 IAGIGGEKAG 10  
 :|||:|  
 DB 469 VGGIGGEAG 478

RESULT 6  
 CCAS\_CYPCA  
 ID \_CCAS\_CYPCA STANDARD; PRT; 1852 AA.  
 AC P22316;  
 DT 01-AUG-1991 (Rel. 19, Created)  
 DT 01-AUG-1991 (Rel. 19, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Dihydropyridine-sensitive L-type, skeletal muscle calcium channel  
 DE alpha-1 subunit.  
 OS *Cyprinus carpio* (Common carp).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;  
 OC Cyprinidae; Cyprinus.  
 OX NCBI\_TaxID=7962;  
 (1)  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Skeletal muscle;  
 RX MEDLINE=91126068; PubMed=1846962;  
 RA Grabner M., Friedrich K., Knaus H.-G., Striessnig J., Scheffauer F.,  
 RA Staedinger R., Koch W.J., Schwartz A., Glossmann H.;  
 RT "Calcium channels from *Cyprinus carpio* skeletal muscle."  
 RL Proc. Natl. Acad. Sci. U.S.A. 88:727-731(1991).  
 CC -!- FUNCTION: Voltage-sensitive calcium channels (VSCC) mediate the  
 CC entry of calcium ions into excitable cells and are also involved  
 CC in a variety of calcium-dependent processes, including muscle  
 CC contraction, gene expression, cell motility, cell division and  
 CC cell death. The isoform alpha-1S gives rise to L-type calcium  
 CC currents. Long-lasting (L-type) calcium channels belong to the  
 CC "high-voltage activated" (HVA) group. They are blocked by  
 CC dihydropyridines (DHP), phenylalkylamines, benzothiazepines, and  
 CC B omega-agatoxin-111A (omega-aga-111A). They are however  
 CC insensitive to omega-conotoxin-GVIA (omega-CTX-GVIA) and omega-  
 CC agatoxin-IVA (omega-Aga-IVA). Calcium channels containing the  
 CC alpha-1S subunit play an important role in excitation-contraction  
 CC coupling in skeletal muscle (by similarity).  
 CC -!- SUBUNIT: VOLTAGE-DEPENDENT CALCIUM CHANNELS ARE MULTISUBUNIT  
 CC COMPLEXES, CONSISTING OF ALPHA-1, ALPHA-2, BETA AND DELTA SUBUNITS  
 CC IN A 1:1:1:1 RATIO. THE CHANNEL ACTIVITY IS DIRECTED BY THE PORE-  
 CC FORMING AND VOLTAGE-SENSITIVE ALPHA-1 SUBUNIT. IN MANY CASES, THIS  
 CC SUBUNIT IS SUFFICIENT TO GENERATE VOLTAGE-SENSITIVE CALCIUM



RA Harrison D., Hoang L., Keagle P., Lumm W., Pothier B., Qiu D.,  
RA Spadafora R., Vicare R., Wang Y., Wierzbowski J., Gibson R.,  
RA Jiwan N., Caruso A., Bush D., Safer H., Patwell D., Prabhakar S.,  
RA McDougall S., Shimer G., Goyal A., Pietrowski S., Church G.M.,  
RA Daniels C.J., Mac J.-I., Rice P., Noelling J., Reeve J.N.,  
RT "Complete genome sequence of Methanobacterium thermoautotrophicum  
RT deltaH: functional analysis and comparative genomics.";  
RL J. Bacteriol. 179:7135-7155(1997).  
CC -!- FUNCTION: CATALYZES THE REVERSIBLE OXIDATION OF CO(2) AND  
CC METHANOFURAN (MFR) TO N-FORMYLMETHANOFURAN (CHO-MFR). CAN ONLY  
CC OXIDISE FORMYLMETHANOFURAN. THIS ENZYME IS OXYGEN-LABILE.  
CC -!- CATALYTIC ACTIVITY: Formylmethanofuran + H(2)O + acceptor = CO(2)  
CC + methanofuran + reduced acceptor.  
CC -!- COFACTOR: Tungsten.  
CC -!- ENZYME REGULATION: Not inactivated by cyanide.  
CC -!- PATHWAY: Methanogenesis; first step.  
CC -!- SUBUNIT: THIS ENZYME IS COMPOSED OF SEVEN SUBUNITS FWDA (65 kDa),  
CC FWDB (53 kDa), FWDC (31 kDa), FWDD (15 kDa), FWDE, FWDF, AND FWDG.  
CC -!- INDUCTION: By growth on tungsten or molybdenum under anaerobic  
CC conditions.  
CC -!- SIMILARITY: Belongs to the fwdC/fwdC family.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; AE000916; AAB86032.1; -;  
DR PIR; H69074; H69074.1;  
DR InterPro; IPR002489; DUF14.  
DR Pfam; PF01493; GXGKG; 1.  
KW Oxidoreductase; Tungsten; Methanogenesis; Repeat; Complete proteome.  
FT FT 7 X 13 AA REPEATS OF [GW]-X-X-M-X-X-G-  
FT DOMAIN 80 213 X-[IL]-X-[IV]-X-G.  
FT REPEAT 80 92 1.  
FT REPEAT 99 111 2.  
FT REPEAT 118 130 3.  
FT REPEAT 144 156 4.  
FT REPEAT 163 175 5.  
FT REPEAT 182 194 6.  
FT REPEAT 201 213 7.  
SQ SEQUENCE 270 AA; 28641 MW; 043A9FFBBA554D36 CRC64;

Query Match 64.5%; Score 40; DB 1; Length 270;  
Best Local Similarity 63.6%; Pred. No. 23;  
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 IAGIGGEXAGG 11  
:|:|:|:|:|  
Db 197 VARVGEMAGG 207

RESULT 8  
FWDC.METTM  
ID FWDC.METTM STANDARD; PRT; 270 AA.  
AC Q59579; O08493;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE Tungsten-containing formylmethanofuran dehydrogenase II subunit C  
DE (EC 1.2.99.5).  
GN FWDC.  
OS Methanobacterium thermoautotrophicum (strain Marburg / DSM 2133).  
OC Archaea; Euryarchaeota; Methanobacteria; Methanobacteriales;  
OC Methanobacteriacea; Methanothermobacter.  
OX NCBI\_TaxID=79929;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=96163477; PubMed=8575452;  
RA Hochheimer A., Schmitz R.A., Thauer R.K., Hedderich R.;

RT "The tungsten formylmethanofuran dehydrogenase from Methanobacterium  
RT thermoautotrophicum contains sequence motifs characteristic for  
RT enzymes containing molybdopterin dinucleotide.";  
RL Eur. J. Biochem. 234:910-920(1995).  
CC -!- FUNCTION: CATALYZES THE REVERSIBLE OXIDATION OF CO(2) AND  
CC METHANOFURAN (MFR) TO N-FORMYLMETHANOFURAN (CHO-MFR). CAN ONLY  
CC OXIDISE FORMYLMETHANOFURAN. THIS ENZYME IS OXYGEN-LABILE.  
CC -!- CATALYTIC ACTIVITY: Formylmethanofuran + H(2)O + acceptor = CO(2)  
CC + methanofuran + reduced acceptor.  
CC -!- COFACTOR: Tungsten.  
CC -!- PATHWAY: Methanogenesis; first step.  
CC -!- SUBUNIT: THIS ENZYME IS COMPOSED OF SEVEN SUBUNITS FWDA (65 kDa),  
CC FWDB (53 kDa), FWDC (31 kDa), FWDD (15 kDa), FWDE, FWDF, AND FWDG.  
CC -!- INDUCTION: By growth on tungsten or molybdenum under anaerobic  
CC conditions.  
CC -!- SIMILARITY: Belongs to the fwdC/fwdC family.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; X87970; CAA61214.1; -;  
DR InterPro; IPR002489; DUF14.  
DR Pfam; PF01493; GXGKG; 1.  
KW Oxidoreductase; Tungsten; Methanogenesis; Repeat.  
FT FT 7 X 13 AA REPEATS OF [GW]-X-X-M-X-X-G-  
FT DOMAIN 80 213 X-[IL]-X-[IV]-X-G.  
FT REPEAT 80 92 1.  
FT REPEAT 99 111 2.  
FT REPEAT 118 130 3.  
FT REPEAT 144 156 4.  
FT REPEAT 163 175 5.  
FT REPEAT 182 194 6.  
FT REPEAT 201 213 7.  
SQ SEQUENCE 270 AA; 28554 MW; E0A369D2ACFEC46F CRC64;

Query Match 64.5%; Score 40; DB 1; Length 270;  
Best Local Similarity 63.6%; Pred. No. 23;  
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 IAGIGGEXAGG 11  
:|:|:|:|:|  
Db 197 VARVGEMAGG 207

RESULT 9  
DNAJ.PORGI  
ID DNAJ.PORGI STANDARD; PRT; 383 AA.  
AC Q9XCAG;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 15-MAR-2004 (Rel. 43, Last annotation update)  
DE Chaperone protein dnaJ (Immunoreactive heat shock protein dnaJ).  
GN DNAJ OR PGI1776.  
OS Porphyromonas gingivalis (Bacteroides gingivalis).  
OC Bacteria; Bacteroidetes; Bacteroides (Class); Bacteroidales;  
OC Porphyromonadaceae; Porphyromonas.  
OX NCBI\_TaxID=837;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=W50;  
RA Ross B.C., Barr I., Patterson M., Agius C., Rothel L., Margetts M.,  
RA Hocking D., Webb E.;  
RT "Porphyromonas gingivalis polypeptides and nucleic acids.";  
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=W83;

```

RX MEDLINE=22829867; PubMed=12949112;
RA Nelson K.E., Fleischmann R.D., DeBoy R.T., Paulsen I.T., Fouts D.E.,
RA Bisen J.A., Dougherty S.C., Dodson R.J., Durkin A.S., Gwinn M.,
RA Haft D.H., Kolonay J.F., Nelson W.C., Mason T., Tallon L., Gray J.,
RA Granger D., Tettelin H., Dong H., Galvin J.L., Duncan M.J.,
RA Dewhirst F.E., Fraser C.M.;
RA "Complete genome sequence of the oral pathogenic bacterium
RT Porphyromonas gingivalis strain W83."
RL J. Bacteriol. 185:5591-5601(2003).
CC -!- FUNCTION: Acts as a co-chaperone. Stimulates, jointly with grpE,
CC the ATPase activity of dnaK (By similarity).
CC -!- COFACTOR: Binds 2 zinc ions per monomer (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: Belongs to the dnaJ family.
CC -!- SIMILARITY: Contains 1 J domain.
CC -!- SIMILARITY: Contains 1 CR domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF145797; AAD39493.1; -.
DR EMBL; AB017178; AAO66777.1; -.
DR HSSP; P08622; 1XBL.
DR TIGR; PGI776; -.
DR InterPro; IPR002939; DnaJ_C.
DR InterPro; IPR001305; DnaJ_CXXCXGKG.
DR InterPro; IPR001623; DnaJ_N.
DR InterPro; IPR008971; HSP40_DnaJ_pdp.
DR InterPro; IPR003095; Hsp_DnaJ.
DR Pfam; PF00226; DnaJ_1.
DR Pfam; PF01556; DnaJ_C; 1.
DR Pfam; PF00684; DnaJ_CXXCXGKG; 1.
DR PRINTS; PR00625; DnaJPROTEIN.
DR SMART; SM00271; DnaJ_1.
DR PROSITE; PS00636; DnaJ_1; 1.
DR PROSITE; PS0076; DnaJ_2; 1.
DR PROSITE; PS00637; DnaJ_CXXCXGKG; FALSE NEG.
DR Chaperone; DNA replication; Heat shock; Repeat; Zinc; Metal-binding;
KW Complete proteome.
FT DOMAIN 6 71 J-DOMAIN.
FT REPEAT 74 124 CXXCXGKG MOTIF.
FT REPEAT 154 161 CXXCXGKG MOTIF.
FT REPEAT 171 178 CXXCXGKG MOTIF.
FT REPEAT 197 204 CXXCXGKG MOTIF.
FT REPEAT 211 218 CXXCXGKG MOTIF.
FT METAL 154 154 ZINC 1 (BY SIMILARITY).
FT METAL 157 157 ZINC 1 (BY SIMILARITY).
FT METAL 171 171 ZINC 2 (BY SIMILARITY).
FT METAL 174 174 ZINC 2 (BY SIMILARITY).
FT METAL 197 197 ZINC 2 (BY SIMILARITY).
FT METAL 200 200 ZINC 2 (BY SIMILARITY).
FT METAL 211 211 ZINC 1 (BY SIMILARITY).
FT METAL 214 214 ZINC 1 (BY SIMILARITY).
SQ SEQUENCE 383 AA; 41106 MW; 449F5020DC7AD596 CRC64;

Query Match 64.5%; Score 40; DB 1; Length 383;
Best Local Similarity 63.6%; Pred. No. 31;
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 2 AGIGGEKAGGF 12
DB 73 AGLGGAAGGF 83

RESULT 10
YCCA CABEL STANDARD; PRT; 175 AA.
ID YCCA CABEL
AC Q23280;


```

```

DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical 19.4 kDa protein ZC395.10 in chromosome III.
GS ZC395.10.
OS Caenorhabditis elegans.
OC Rukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Connell M.;
RL Submitted (NOV-1994) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: Belongs to the p23 / wos2 family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U13642; AAG00038.1; -.
DR PIR; T27543; T27543.
DR WormPep; ZC395.10; CE01436.
DR InterPro; IPR008978; HSP20_chap.
KW Hypothetical protein.
FT DOMAIN 145 175 ASP/GLU-RICH.
FT DOMAIN 165 168 POLY-GLU.
SQ SEQUENCE 175 AA; 19431 MW; D5C136F30446E37A CRC64;

Query Match 62.9%; Score 39; DB 1; Length 175;
Best Local Similarity 72.7%; Pred. No. 22;
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AGIGGEKAGGF 12
DB 119 AGIGGMANGF 129

RESULT 11
REF1 LEPIN STANDARD; PRT; 354 AA.
AC Q8F986;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DE Peptide chain release factor 1 (RF-1).
DE PRFA OR LA0114.
OS Leptospira interrogans.
OC Bacteria; Spirochaetes; Spirochaetales; Leptospiraceae; Leptospira.
OX NCBI_TaxID=173;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=56601 / Serogroup Icterohaemorrhagiae / Serovar lai;
RX MEDLINE=22598143; PubMed=12712204;
RA Ren S.-X., Fu G., Jiang X.-G., Zeng R., Miao Y.-G., Xu H.,
RA Zhang Y.-X., Xiong H., Lu G., Lu L.-F., Jiang H.-Q., Jia J., Tu Y.-F.,
RA Jiang J.-X., Gu W.-Y., Zhang Y.-Q., Cai Z., Sheng H.-H., Yin H.-F.,
RA Zhang Y., Zhu G.-F., Wan M., Huang H.-L., Qian Z., Wang S.-Y., Ma W.,
RA Yao Z.-J., Shen Y., Qiang B.-Q., Xia O.-C., Guo X.-K., Dauchin A.,
RA Saint Girons I., Somerville R.L., Wen Y.-M., Shi M.-H., Chen Z.,
RA Xu J.-G., Zhao G.-P.;
RP SEQUENCE FROM N.A.
RC STRAIN=56601 / Serogroup Icterohaemorrhagiae / Serovar lai;
RX MEDLINE=22598143; PubMed=12712204;
RA Ren S.-X., Fu G., Jiang X.-G., Zeng R., Miao Y.-G., Xu H.,
RA Zhang Y.-X., Xiong H., Lu G., Lu L.-F., Jiang H.-Q., Jia J., Tu Y.-F.,
RA Jiang J.-X., Gu W.-Y., Zhang Y.-Q., Cai Z., Sheng H.-H., Yin H.-F.,
RA Zhang Y., Zhu G.-F., Wan M., Huang H.-L., Qian Z., Wang S.-Y., Ma W.,
RA Yao Z.-J., Shen Y., Qiang B.-Q., Xia O.-C., Guo X.-K., Dauchin A.,
RA Saint Girons I., Somerville R.L., Wen Y.-M., Shi M.-H., Chen Z.,
RA Xu J.-G., Zhao G.-P.;
RT "Unique physiological and pathogenic features of Leptospira
RT interrogans revealed by whole-genome sequencing."
RL Nature 422:888-893(2003).
CC -!- FUNCTION: Peptide chain release factor 1 directs the termination
CC of translation in response to the peptide chain termination codons
CC UAG and UAA (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: Belongs to the prokaryotic/mitochondrial release


```

factor family.

-----  
 This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 -----

EMBL; AE011202; AAN47313.1; --  
 HAMAP; MF 00093; -; 1.  
 InterPro; IPR005139; PCRF.  
 InterPro; IPR00352; Pep\_rel\_factor\_1.  
 InterPro; IPR004373; PfIA.  
 Pfam; PF03462; PCRF; 1.  
 Pfam; PF00472; RF-1; 1.  
 TIGRFAMs; TIGR00019; drfa; 1.  
 PROSITE; PS00745; RE\_PROK\_I; 1.  
 Protein biosynthesis; Complete proteome.

KW SEQUENCE 354 AA; 40099 MW; 779A91ACCC997A59 CRC64;  
 SQ

Query Match 62.9%; Score 39; DB 1; Length 354;  
 Best Local Similarity 72.7%; Pred. No. 42;  
 Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 2 AGIGGEKAGGF 12  
 |||||  
 Db 113 AGTGGEEAGLF 123

RESULT 12

CSP\_PLACL ID - CSP\_PLACL STANDARD; PRT; 378 AA.  
 AC P08675;  
 DT 01-JAN-1988 (Rel. 06, Created)  
 DT 01-JAN-1988 (Rel. 06, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE Circumsporozoite protein precursor (CS).  
 OS Plasmodium cynomolgi (strain London).  
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
 OX NCBI\_TaxID=5831;  
 RN [1]

SEQUENCE FROM N.A.  
 RP MEDLINE=87102878; PubMed=3802196;  
 RA Galinski M.R., Arnot D.E., Cochran A.H., Barnwell J.W.,  
 RA Nussenzweig R.S., Enea V.;  
 RT "The circumsporozoite gene of the Plasmodium cynomolgi complex.";  
 RL Cell 48:311-319(1987).  
 CC -!- MISCELLANEOUS: The circumsporozoite protein is the immunodominant  
 CC surface antigen on the sporozoite (the infective stage of the  
 CC malaria parasite that is transmitted from the mosquito to the  
 CC vertebrate host).  
 CC -!- MISCELLANEOUS: The C-terminal region is probably used for  
 CC anchoring the protein to the cell membrane. The repeat sequences  
 CC would be the surface antigen of the organism.  
 CC -!- SIMILARITY: Contains 1 TSP type-1 domain.  
 -----

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 -----

EMBL; M15101; AAA29537.1; --  
 InterPro; IPR003067; Crmsprzoite.  
 InterPro; IPR000884; TSP1.  
 Pfam; PF00090; tsp\_1; 1.  
 PRINTS; PR01303; CRMSPRZOITE.  
 SMART; SM00209; TSP1; 1.  
 PROSITE; PS50092; TSP1; 1.

Malaria; Sporozoite; Repeat; Signal.  
 KW SIGNAL 1 19 PROBABLE.  
 FT CHAIN 20 378 CIRCUMSPOROZOITE PROTEIN.  
 FT DOMAIN 99 211 18 X 6 AA TANDEM REPEATS OF D-G-A-R-A-  
 FT [EA].  
 FT REPEAT 99 104 1-1.  
 FT REPEAT 105 110 1-2.  
 FT REPEAT 111 116 1-3.  
 FT REPEAT 117 122 1-4.  
 FT REPEAT 123 128 1-5.  
 FT REPEAT 129 134 1-6.  
 FT REPEAT 135 140 1-7.  
 FT REPEAT 141 146 1-8.  
 FT REPEAT 147 152 1-9.  
 FT REPEAT 153 158 1-10.  
 FT REPEAT 159 164 1-11.  
 FT REPEAT 165 170 1-12.  
 FT REPEAT 171 176 1-13.  
 FT REPEAT 177 182 1-14.  
 FT REPEAT 183 188 1-15.  
 FT REPEAT 189 194 1-16.  
 FT REPEAT 195 200 1-17.  
 FT REPEAT 201 206 1-18.  
 FT DOMAIN 212 277 6 X 11 AA TANDEM REPEATS OF G-N-[QR] -  
 FT [AE]-G-G-Q-A-G-A-G.  
 FT REPEAT 212 222 2-1.  
 FT REPEAT 223 233 2-2.  
 FT REPEAT 234 244 2-3.  
 FT REPEAT 245 255 2-4.  
 FT REPEAT 256 266 2-5.  
 FT REPEAT 267 277 2-6.  
 FT DOMAIN 304 356 TSP TYPE-1.  
 SQ SEQUENCE 378 AA; 37462 MW; 8295A913C36420C5 CRC64;

Query Match 62.9%; Score 39; DB 1; Length 378;  
 Best Local Similarity 70.0%; Pred. No. 45;  
 Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 2 AGIGGEKAGG 11  
 |||||  
 Db 263 AGAGGNRAGG 272

RESULT 13

HXAA\_MOUSE ID - HXAA\_MOUSE STANDARD; PRT; 399 AA.  
 AC P31310;  
 DT 01-JUL-1993 (Rel. 26, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Homeobox protein Hox-A10 (Hox-1.8).  
 GN HOXA10 OR HOXA-10 OR HOX-1.8.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]

SEQUENCE FROM N.A.  
 RP STRAIN=CD-1; TISSUE=Kidney;  
 RX MEDLINE=95166244; PubMed=7862151;  
 RA Benson G.V., Nguyen T.-H.E., Maas R.L.;  
 RT "The expression pattern of the murine Hoxa-10 gene and the sequence  
 RT recognition of its homeodomain reveal specific properties of  
 RT Abdominal B-like genes.";  
 RL Mol. Cell. Biol. 15:1591-1601(1995).  
 RN [2]

SEQUENCE OF 325-384 FROM N.A.  
 RX MEDLINE=92073356; PubMed=1683707;  
 RA Singh G., Kaur S., Stock J.L., Jenkins N.A., Gilbert D.J.,  
 RA Copeland N.G., Potter S.S.;  
 RT "Identification of 10 murine homeobox genes.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 88:10706-10710(1991).  
 RN [3]

```

RP SEQUENCE OF 346-370 FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Spleen;
RX MEDLINE=92073357; PubMed=1720547;
RA Murtha M.T., Leckman J.F., Ruddle F.H.;
RL "Detection of homeobox genes in development and evolution.";
RT Proc. Natl. Acad. Sci. U.S.A. 88:10711-10715(1991).
CC -!- FUNCTION: SEQUENCE-SPECIFIC TRANSCRIPTION FACTOR WHICH IS PART OF
CC A DEVELOPMENTAL REGULATORY SYSTEM THAT PROVIDES CELLS WITH
CC SPECIFIC POSITIONAL IDENTITIES ON THE ANTERIOR-POSTERIOR AXIS.
CC BINDS TO THE DNA SEQUENCE 5'-AA[AT]TTTATTAC-3'.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=1;
CC IsoId=P31310-1; Sequence=Displayed;
CC Name=2;
CC IsoId=P31310-2; Sequence=VSP_002386, VSP_002387;
CC -!- TISSUE SPECIFICITY: EXPRESSED IN THE DEVELOPING LIMB BUD WHERE IT
CC IS RESTRICTED TO THE MESENCHYME ALONG THE PROXIMAL-DISTAL AXIS.
CC ALSO FOUND IN DEVELOPING GUT AND UROGENITAL TRACT. IN ADULT
CC TISSUE, BOTH FORMS FOUND IN KIDNEY BUT ONLY ISOFORM 1 IS EXPRESSED
CC IN SKELETAL MUSCLE.
CC -!- DEVELOPMENTAL STAGE: EMBRYONIC EXPRESSION INCREASES FROM DAY 9 TO
CC DAY 12 AND THEN DECLINES TO DAY 15.
CC -!- SIMILARITY: Belongs to the Abd-B homeobox family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L08757; AAA67125.1; -.
DR EMBL; M81659; AAA63312.1; -.
DR PIR; I49754; I49754.
DR HSP; P02833; 9ANT.
DR TRANSFAC; T01715; -.
DR MGD; MGI:96171; Hoxa10.
DR InterPro; IPR001356; Homeobox.
DR Pfam; PF00046; homeobox; 1.
DR PRINTS; PR00024; HOMEBOX.
DR ProDom; PD00010; Homeobox; 1.
DR SMART; SM00389; HOX; 1.
DR PROSITE; PS00027; HOMEBOX_1; 1.
DR PROSITE; PS00071; HOMEBOX_2; 1.
KW Homeobox; DNA-binding; Developmental protein; Nuclear protein;
KW Transcription regulation; Alternative splicing.
FT DOMAIN 26 38 GLY-RICH.
FT DOMAIN 77 90 POLY-GLY.
FT DOMAIN 113 144 GLN/PRO-RICH.
FT DOMAIN 215 222 POLY-GLY.
FT DOMAIN 271 274 POLY-GLY.
FT DOMAIN 325 384 HOMEBOX.
FT VARSPLIC 1 305 MISSING (in isoform 2).
FT VARSPLIC 306 308 DSL->MCQ (in isoform 2).
FT VARSPLIC 309 308 /FTRGVSP_002387.
SQ SEQUENCE 399 AA; 14145 MW; 7529624FC6057042 CRC64;
Query Match 62.9%; Score 39; DB 1; Length 399;
Best Local Similarity 70.0%; Pred. No. 47;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 2 AGIGEXKAGG 11
Db 25 AGVGGSGAGG 34
RESULT 14
FUR2_DROME
ID FUR2_DROME STANDARD; PRT; 1679 AA.

```

```

AC P30432; Q24301;
DT 01-APR-1993 (Rel. 25, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Furin-like protease 2 precursor (EC 3.4.21.75) (Furin 2).
GN FUR2 OR CG18734/CG4235.
OS Drosophila melanogaster (fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Oregon-R, Tuebingen, and Iso-1;
RX MEDLINE=92381036; PubMed=1512259;
RA Roebroek A.J.M., Creemers J.W.M., Pauli I.G.L., Kurzik-Dunke U.,
RA van de Ven W.J.M.;
RA "Cloning and characterization of Dfurin2, a subtilisin-like
RT proprotein processing enzyme of Drosophila melanogaster with multiple
RT repeats of a cysteine motif.";
RL J. Biol. Chem. 267:17208-17215(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Iso-1;
RX MEDLINE=95186060; PubMed=7880443;
RA Roebroek A.J.M., Ayoubi T.A.Y., Creemers J.W.M., Pauli I.G.L.,
RA van de Ven W.J.M.;
RA "The Dfur2 gene of Drosophila melanogaster: genetic organization,
RT expression during embryogenesis, and pro-protein processing activity
RT of its translational product Dfurin2.";
RL DNA Cell Biol. 14:223-234(1995).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkely;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yeung R.G., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazer V.G., Champ M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu P., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrier S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Laslo P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,
RA Liu X., Mates J., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reibert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhou S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).

```

FT	CARBOHYD	1181	1181	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	1274	1274	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	1277	1277	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	1439	1439	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	152	153	MISSING (IN REF. 1).	
FT	CONFLICT	177	177	V -> F (IN REF. 1).	
FT	CONFLICT	213	213	V -> VDQL (IN REF. 1).	
SEQ	SEQUENCE	1679 AA,	183369 MW,	3F9E749F0B021CF6	CRC64;

Query Match 62.9%; Score 39; DB 1; Length 1679;  
Best Local Similarity 60.0%; Pred. No. 1.7e+02;  
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY	1	IAGIGGEKAG	10
DB	176	LVGLGGERAG	185

: : |||: |||  
: : |||: |||

ID MTRC\_METAC STANDARD; PRT: 267 AA.

Q8TU01;  
15-MAR-2004 (Rel. 43, Created)  
15-MAR-2004 (Rel. 43, Last sequence update)  
15-MAR-2004 (Rel. 43, Last annotation update)  
Tetrahydromethanopterin S-methyltransferase subunit C (EC 2.1.1.86)  
(N5-methyltetrahydromethanopterin--coenzyme M methyltransferase subunit C)  
MTRC OR MA0274.

Methanosarcina acetivorans.  
Archaea; Euryarchaeota; Methanomicrobia; Methanosarcinales;  
Methanosarcinaceae; Methanosarcina.  
NCBI\_TaxID=2214;  
[1]  
SEQUENCE FROM N.A.  
STRAIN=CZA / ATCC 35395 / DSM 2834;  
MEDLINE=21929760; PubMed=11932238;  
Galagan J.E., Nussbaum C., Roy A., Endrizzi M.G., Macdonald P.,  
FitzHugh W., Calvo S., Engels R., Smirnov S., Anurov D., Brown A.,  
Allen N., Naylor J., Stange-Thomann N., DeAtellano K., Johnson R.,  
Linton L., McEwan P., McKernan K., Talamas J., Tirrelli A., Ye W.,  
Zimmer A., Barber R.D., Cann I., Graham D.E., Grahame D.A., Guss A.M.,  
Hedderich R., Ingram-Smith C., Kuettnner H.C., Krzycki J.A.,  
Leigh J.A., Li W., Liu J., Mukhopadhyay B., Reeve J.N., Smith K.,  
Springer T.A., Umayam L.A., White O., White R.H., de Macario E.C.,  
Ferry J.G., Jarrell K.F., Jing H., Macario A.J.L., Paulsen I.,  
Pritchett M., Sowers K.R., Swanson R.V., Zinder S.H., Lander E.,  
Metcalf W.W., Birren B.;  
The genome of Methanosarcina acetivorans reveals extensive metabolic  
and physiological diversity.";  
Genome Res. 12:532-542(2002).

-!- FUNCTION: Part of a complex that catalyzes the formation of  
methyl-coenzyme M and tetrahydromethanopterin from coenzyme M and  
methyl-tetrahydromethanopterin. This is an energy-conserving,  
sodium-ion translocating step (By similarity).  
-!- CATALYTIC ACTIVITY: 5-methyl-5,6,7,8-tetrahydromethanopterin + 2-  
mercaptoethanesulfonate = 5,6,7,8-tetrahydromethanopterin + 2-  
(methylthio)ethanesulfonate.  
-!- PATHWAY: Methanogenesis from carbon dioxide; sixth step.  
-!- SUBUNIT: The complex is composed of 8 subunits; mtrA, mtrB, mtrC,  
mtrD, mtrE, mtrF, mtrG and mtrH (By similarity).  
-!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).  
-!- SIMILARITY: Belongs to the mtrC family.

This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL: AF010686; AM03727.1; -

DR HAMAP: MF\_01096: -: 1.  
DR InterPro: IPR005865; THM\_met\_transc.  
DR Pfam: PF04211; MtrC; 1.  
KW Methanogenesis; One-carbon metabolism; Transferase; Methyltransferase;  
KW Transmembrane; Complete proteome.  
FT TRANSMEM 19 39 POTENTIAL.  
FT TRANSMEM 40 60 POTENTIAL.  
FT TRANSMEM 75 95 POTENTIAL.  
FT TRANSMEM 96 116 POTENTIAL.  
FT TRANSMEM 131 151 POTENTIAL.  
FT TRANSMEM 162 182 POTENTIAL.  
FT TRANSMEM 221 241 POTENTIAL.  
SQ SEQUENCE 267 AA; 26950 MW; FD258F13BE4D92ED CRC64;  
  
Query Match 61.3%; Score 38; DB 1; Length 267;  
Best Local Similarity 70.0%; Pred. No. 47;  
Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
  
QY 3 GIGGEKAGGF 12  
DB 5 GAGGEAKGGF 14

Search completed: September 18, 2004, 04:26:25  
Job time : 9.94118 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 18, 2004, 04:00:08 : Search time 42.5882 Seconds  
(without alignments)  
88.903 Million cell updates/sec

Title: US-10-615-959-43  
Perfect score: 62  
Sequence: 1 IAGIGGEKAGGF 12

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

- SPTREMBL 25:\*
- 1: sp\_archaea:\*
  - 2: sp\_bacteria:\*
  - 3: sp\_fungi:\*
  - 4: sp\_human:\*
  - 5: sp\_invertebrate:\*
  - 6: sp\_mammal:\*
  - 7: sp\_mhc:\*
  - 8: sp\_organelle:\*
  - 9: sp\_phage:\*
  - 10: sp\_plant:\*
  - 11: sp\_rodent:\*
  - 12: sp\_virus:\*
  - 13: sp\_vertebrate:\*
  - 14: sp\_unclassified:\*
  - 15: sp\_rvirus:\*
  - 16: sp\_bacteriap:\*
  - 17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	62	100.0	1163	4 Q8N6U4	Q8N6U4 homo sapien
2	58	93.5	338	11 Q8BUJ6	Q8BUJ6 mus musculus
3	58	93.5	1222	11 Q8K173	Q8K173 mus musculus
4	58	93.5	1464	11 Q8BKY2	Q8BKY2 mus musculus
5	58	93.5	1464	11 Q8BLW4	Q8BLW4 mus musculus
6	54	87.1	1464	11 Q7WT32	Q7WT32 mus musculus
7	44	71.0	487	16 Q53174	Q53174 mycobacteri
8	44	71.0	487	16 Q7TYK1	Q7TYK1 mycobacteri
9	42	67.7	209	16 Q8UAC0	Q8UAC0 agrobacteri
10	42	67.7	1755	10 Q7XLB7	Q7XLB7 oryza sativ
11	41	66.1	208	10 Q8L418	Q8L418 oryza sativ
12	41	66.1	379	16 Q92Q09	Q92Q09 rhizobium m
13	41	66.1	383	17 Q9YFP0	Q9YFP0 aeropyrum p
14	41	66.1	460	2 Q8RSI9	Q8RSI9 gamma-prote
15	41	66.1	974	5 Q27376	Q27376 caenorhabdi
16	41	66.1	974	5 Q96519	Q96519 caenorhabdi

17	40	64.5	136	11 Q9D3C5	Q9D3C5 mus musculu
18	40	64.5	156	10 Q9FRJ4	Q9FRJ4 oryza sativ
19	40	64.5	156	10 Q7XCM3	Q7XCM3 oryza sativ
20	40	64.5	298	16 Q93038	Q93038 rhizobium m
21	40	64.5	412	2 Q50371	Q50371 methylphil
22	40	64.5	438	16 Q25708	Q25708 helicobacte
23	40	64.5	438	16 Q9ZM72	Q9ZM72 helicobacte
24	40	64.5	548	3 Q9P3K6	Q9P3K6 neuropeptid
25	40	64.5	569	11 Q7TRJ5	Q7TRJ5 rattus norv
26	40	64.5	696	10 Q7XPP9	Q7XPP9 oryza sativ
27	40	64.5	916	16 Q87ZN1	Q87ZN1 pseudomonas
28	39	62.9	87	10 Q36779	Q36779 daucus caro
29	39	62.9	90	10 Q9FTT5	Q9FTT5 oryza sativ
30	39	62.9	147	16 Q8PFE1	Q8PFE1 xanthomonas
31	39	62.9	150	16 Q8P3W1	Q8P3W1 xanthomonas
32	39	62.9	163	10 Q9M238	Q9M238 arabidopsis
33	39	62.9	189	10 Q94J02	Q94J02 oryza sativ
34	39	62.9	232	10 Q8H5B2	Q8H5B2 oryza sativ
35	39	62.9	237	10 Q7XXK9	Q7XXK9 oryza sativ
36	39	62.9	240	10 Q8LC37	Q8LC37 arabidopsis
37	39	62.9	240	10 Q94A32	Q94A32 arabidopsis
38	39	62.9	372	16 Q87I01	Q87I01 vibrio para
39	39	62.9	386	16 Q8VJY7	Q8VJY7 mycobacteri
40	39	62.9	390	10 Q7XR48	Q7XR48 oryza sativ
41	39	62.9	400	10 Q8GIL0	Q8GIL0 oryza sativ
42	39	62.9	460	2 Q7WY41	Q7WY41 pseudomonas
43	39	62.9	575	10 Q9LIR6	Q9LIR6 arabidopsis
44	39	62.9	605	16 Q97NH2	Q97NH2 streptococc
45	39	62.9	605	16 Q8CY83	Q8CY83 streptococc

ALIGNMENTS

RESULT 1

Q8N6U4

ID Q8N6U4 PRELIMINARY; PRT; 1163 AA.

AC Q8N6U4; 01-OCT-2002 (Tremblrel. 22, Created)

DT 01-OCT-2002 (Tremblrel. 22, Last sequence update)

DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)

DE Similar to collagen, type III, alpha 1 (Ehlers-Danlos syndrome type IV, autosomal dominant).

DE IV, autosomal dominant).

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI\_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Liver;

RA Strausberg R.;

RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.

DR EMBL; BC028178; AAH28178.1; -

DR GO; GO:0005581; C:collagen; IEA.

DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.

DR InterPro; IPR008160; Collagen.

DR InterPro; IPR002181; Fibrinogen\_C.

DR InterPro; IPR000885; Fib.collagen\_C.

DR Pfam; PF01410; COLFI; 1.

DR Pfam; PF01391; Collagen; 13.

DR	ProDom; PD002078; Fib.collagen\_C; 1.
DR	SMART; SM00038; COLFI; 1.
DR	SMART; SM00214; VWC; 1.
DR	PROSITE; PS01208; VWC; 1.
DR	PROSITE; PS01084; VWC\_2; 1.
DR	Collagen.
DR	SEQUENCE 1163 AA; 111899 MW; 9E0C6BE1B94D6357 CRC64;

Query Match 100.0%; Score 62; DB 4; Length 1163;  
Best Local Similarity 100.0%; Pred. No. 0.24; Indels 0; Gaps 0;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

02

DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Collagen alpha 1.  
 GN COL3A1.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Aorta, and Vein;  
 RX MEDLINE=22354683; PubMed=12466851;  
 RA The FANTOM Consortium,  
 RA the RIKEN Genome Exploration Research Group Phase I & II Team;  
 RT "Analysis of the mouse transcriptome based on functional annotation of  
 RT 60,770 full-length cDNAs."  
 RL Nature 420:563-573(2002).  
 DR EMBL; AK041115; BAC30826.1; -.  
 DR MGD; MGI:188453; Col3a1.  
 DR GO; GO:0005581; C:collagen; IEA.  
 DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.  
 DR InterPro; IPR008161; Clg\_helix.  
 DR InterPro; IPR008160; Collagen.  
 DR InterPro; IPR002181; Fibrinogen\_C.  
 DR InterPro; IPR000885; Fib\_collagen\_C.  
 DR InterPro; IPR001007; VWF\_C.  
 DR Pfam; PF01410; COLFI; 1.  
 DR ProDom; PD000007; Clg\_helix; 1.  
 DR ProDom; PD002078; Fib\_collagen\_C; 1.  
 DR SMART; SM00038; COLFI; 1.  
 DR SMART; SM00214; WVC; 1.  
 DR PROSITE; PS01208; WVC1; 1.  
 DR PROSITE; PS0184; WVC2; 1.  
 SQ SEQUENCE 1464 AA; 136939 MW; 91F3246D90818449 CRC64;  
 Query Match 93.5%; Score 58; DB 11; Length 1464;  
 Best Local Similarity 83.3%; Pred. No. 1.3;  
 Matches 10; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
 OY 1 IAGIGGKAGGF 12  
 DB 1203 IAGVGGEKSGGF 1214  
 RESULT 6  
 Q7TT32 PRELIMINARY; PRT; 1464 AA.  
 ID Q7TT32  
 AC Q7TT32  
 DT 01-OCT-2003 (TrEMBLrel. 25, Created)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)  
 DE Hypothetical protein.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Brain;  
 RX MEDLINE=22388257; PubMed=12477932;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Datchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Frange C.,  
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,

RA Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,  
 RA Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,  
 RA Jones S.J., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length human  
 RT and mouse cDNA sequences."  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Brain;  
 RX Strausberg R.;  
 RL Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; BC052398; AAH52398.1; -.  
 KW Hypothetical protein.  
 SQ SEQUENCE 1464 AA; 138971 MW; 2B38BC27AF21590B CRC64;  
 Query Match 87.1%; Score 54; DB 11; Length 1464;  
 Best Local Similarity 75.0%; Pred. No. 5.9;  
 Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
 OY 1 IAGIGGKAGGF 12  
 DB 1203 IAGVGGEKSGGF 1214  
 RESULT 7  
 Q53174 PRELIMINARY; PRT; 487 AA.  
 ID Q53174  
 AC Q53174  
 DT 01-JUN-1998 (TrEMBLrel. 06, Created)  
 DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Polyglutamate synthase.  
 GN FOLC OR RV2447C OR MT008.03C OR MT2523.  
 OS Mycobacterium tuberculosis.  
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
 OX NCBI\_TaxID=1773;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=H37Rv;  
 RX MEDLINE=96295987; PubMed=9634230;  
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
 RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,  
 RA Badcock K., Basham D., Brown D., Brown S., Chillingworth T., Connor R.,  
 RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,  
 RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,  
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
 RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,  
 RA Rastall J.E., Taylor K., Whitehead S., Barrell B.G.;  
 RT "Deciphering the biology of Mycobacterium tuberculosis from the  
 RT complete genome sequence."  
 RL Nature 393:537-544(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CDC 1551 / Oshkosh;  
 RX Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,  
 RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,  
 RA Kolonay J.F., Nelson W.C., Unayam L.A., Ermolaeva M.D., Salzberg S.L.,  
 RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,  
 RA Bishai W.;  
 RT "Whole genome comparison of Mycobacterium tuberculosis clinical and  
 RT laboratory strains."  
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AL021246; CAAL6024.1; -.  
 DR EMBL; AE007089; AAK46822.1; -.  
 DR PIR; F70863; F70863.  
 DR HSSP; P15925; 1FGS.  
 DR TIGR; MT2523; -.  
 DR TuberculList; RV2447c; -.  
 DR GO; GO:0005524; F:ATP binding; IEA.

```

DR GO: GO:0016874; F.ligase activity; IEA.
DR GO: GO:0004326; Fictetranhydrofolylpolyglutamate synthase activity; IEA.
DR GO: GO:0009058; P-biosynthesis; IEA.
DR GO: GO:0009396; P-folic acid and derivative biosynthesis; IEA.
DR InterPro: IPR001645; Fpolylg synthase.
DR InterPro: IPR000713; Mur_ligase.
DR InterPro: IPR004101; Mur_ligase_C.
DR Pfam: PF01225; Mur_ligase; 1.
DR Pfam: PF02875; Mur_ligase_C; 1.
DR TRIPFAM: TIGR01499; folC; 1.
DR PROSITE: PS01012; FOLYLPOLYGLU_SYNT_2; 1.
KW Complete proteome.
SQ SEQUENCE 487 AA; 50779 MW; 77D3A4E3D2DC59D0 CRC64;

Query Match 71.0%; Score 44; DB 16; Length 487;
Best Local Similarity 90.0%; Pred. No. 78;
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 IAGIGGEKAG 10
Db 209 IAGIAGEKAG 218

RESULT 8
Q7TYK1 PRELIMINARY; PRT; 487 AA.
AC Q7TYK1;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Probable folylpolyglutamate synthase protein FOLC
DE (Folylpoly-gamma-glutamate synthetase) (FPGS) (EC 6.3.2.17).
GN FOLC OR MB2474C.
OS Mycobacterium bovis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1765;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AF2122/97;
RX MEDLINE=22709107; PubMed=12788972;
RA Garner T., Eiglmeier K., Camus J.-C., Medina N., Mansoor H.,
RA Pryor M., Duthoy S., Grondin S., Lacroix C., Monsemp C., Simon S.,
RA Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,
RA Parkhill J., Barrall B.G., Cole S.T., Gordon S.V., Hewinson R.G.;
RT "The complete genome sequence of Mycobacterium bovis."
RL Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882(2003).
DR EMBL: BX248342; CAD97335.1; -.
KW Ligase; Complete proteome.
SQ SEQUENCE 487 AA; 50779 MW; 77D3A4E3D2DC59D0 CRC64;

Query Match 71.0%; Score 44; DB 16; Length 487;
Best Local Similarity 90.0%; Pred. No. 78;
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 IAGIGGEKAG 10
Db 209 IAGIAGEKAG 218

RESULT 9
Q8UAC0 PRELIMINARY; PRT; 209 AA.
AC Q8UAC0;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Arylesterase.
GN ADA OR ATU3454 OR AGR 1.2749.
OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Rhizobium/Agrobacterium group; Agrobacterium.
OX NCBI_TaxID=176299;

```

```

RN RP SEQUENCE FROM N.A.
RX MEDLINE=21608550; PubMed=11743193;
RA Wood D.W., Setubal J.C., Kaul R., Monks D.E., Kitajima J.P.,
RA Okura V.K., Zhou Y., Chen L., Wood G.E., Almeida N.F. Jr., Woo L.,
RA Chen Y., Paulsen I.T., Eisen J.A., Karp P.D., Bovee D. Sr.,
RA Chapman P., Clendenning J., Deatherage G., Gillet W., Grant C.,
RA Kutayin T., Levy R., Li M.-J., McClelland E., Palmieri A.,
RA Raymond C., Rouse G., Saenphimmachak C., Wu Z., Romero P., Gordon D.,
RA Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Perry M.,
RA Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,
RA Chumley P., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,
RA Nester E.W.;
RT "The genome of the natural genetic engineer Agrobacterium tumefaciens
RT C58."
RL Science 294:2317-2323(2001).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=21608551; PubMed=11743194;
RA Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M.,
RA Quorllo B., Goldman B.S., Cao Y., Askenazi M., Helling C., Mullin L.,
RA Houmlel K., Gordon J., Vaudin M., Tatchouk O., Epp A., Liu F.,
RA Wollam C., Allinger M., Doughty D., Scott C., Lappas C., Markelz B.,
RA Flanagan C., Crowell C., Gurson J., Lomo C., Sear C., Strub G.,
RA Cielo C., Slater S.;
RT "Genome sequence of the plant pathogen and biotechnology agent
RT Agrobacterium tumefaciens C58."
RL Science 294:2323-2328(2001).
DR EMBL: AE009275; AAL44267.1; -.
DR EMBL: AE008336; AAK89941.1; -.
DR PIR: AE2981; AE2981.
DR PIR: C98302; C98302.
DR GO: GO:0003824; F:catalytic activity; IEA.
DR InterPro: IPR001087; Lipase_GDSL.
DR Pfam: PF00657; Lipase_GDSL; 1.
KW Complete proteome.
SQ SEQUENCE 209 AA; 21865 MW; B09845588049D0C6 CRC64;

Query Match 67.7%; Score 42; DB 16; Length 209;
Best Local Similarity 72.7%; Pred. No. 68;
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 IAGIGGEKAG 11
Db 137 VAGPGEPPAGG 147

RESULT 10
Q7XLB7 PRELIMINARY; PRT; 1755 AA.
AC Q7XLB7;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE OSJNBA0011K22.12 protein.
GN OSJNBA0011K22.12
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RA Fu G., Wang S.Y., Ren S.X., Lv G., Lin W., Gu W.Q., Zhu G.F., Tu Y.F.,
RA Jia J., Yin H.F., Zhang Y., Cai Z., Chen J., Kang H., Chen X.Y.,
RA Shao Y., Sun Y., Hu Q.P., Zhang X.L., Zhang W., Wang L.J., Ding C.W.,
RA Huang H.H., Gu J.L., Chen S.T., Ni L., Zhu F.H., Han B., Feng Q.,
RA Huang Y.C., Li Y., Zhu J.J., Zhao Q., Hu X., Liu Y.L., Mu J., Yu Z.,
RA Chen T., Fan D.L., Weng Q.J., Zhang L., Lu Y.Q., Yu S.L., Liu X.H.,
RA Lu T.T., Zhang Y.J., Lu Y., Li C., Li T., Zhang Y., Hu H., Jia P.X.,
RA Qian Y.M., Ying K., Zhou B., Chen Z.H., Hao P., Zhang L., Wu M.,
RA Zhang R.Q., Guan J.P., Hong G.F.;
RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.

```

```
DR EMBL; AL731644; CAB05230.1; -.
SQ SEQUENCE 1755 AA; 198541 MW; 9BC452826914DB6 CRC64;

Query Match
Best Local Similarity 67.7%; Score 42; DB 10; Length 1755;
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 1 IAGIGGKAGGF 12
   ||| ||| |||
Db 232 IAGSGDKGGGF 243

RESULT 11
Q8L418
ID Q8L418 PRELIMINARY; PRT; 208 AA.
AC Q8L418;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE P0551A11.6 protein (OJ116 C07.6 protein).
GN P0551A11.6 OR OJ116 C07.6.
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzeae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Nipponbare;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 1, PAC
RT clone:P0551A11.";
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Nipponbare;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 1, BAC
RT clone:OJ116 C07.";
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP003934; BAB92814.1; -.
DR EMBL; AP004253; BAC00689.1; -.
DR Gramene; Q8L418; -.
SQ SEQUENCE 208 AA; 20670 MW; C4D641DBA4D55B80 CRC64;

Query Match
Best Local Similarity 66.1%; Score 41; DB 10; Length 208;
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Qy 1 IAGIGGKAGG 11
   ||| ||| |||
Db 63 VGVGVGREGG 73

RESULT 12
Q92Q09
ID Q92Q09 PRELIMINARY; PRT; 379 AA.
AC Q92Q09;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Probable aminomethyltransferase (Glycine cleavage system T protein)
DE (EC 2.1.2.10).
GN GCVT OR R01549 OR SMC02047.
OS Rhizobium meliloti (Sinorhizobium meliloti).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Sinorhizobium/Ensifer group; Sinorhizobium.
OX NCBI_TaxID=382;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1021;
RX MEDLINE=21396507; PubMed=11481430;
RA Capela D., Barloy-Hubler F., Gouzy J., Bothe G., Ampe F., Batut J.,
```

```
RA Boistard P., Becker A., Boutry M., Cadieu E., Dreano S., Gloux S.,
RA Godrie T., Goffeau A., Kahn D., Kiss E., Leilaure V., Masuy D.,
RA Pohl T., Portetelle D., Puehler A., Purnelle B., Rampsperger U.,
RA Renard C., Thebault P., Vandenbol M., Weidner S., Galibert F.;
RT "Analysis of the chromosome sequence of the legume symbiont
RT Sinorhizobium meliloti strain 1021.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882(2001).
DR EMBL; AL591787; CAC46128.1; -.
DR GO; GO:0004047; F:aminomethyltransferase activity; IEA.
DR GO; GO:0004374; F:glycine cleavage system; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR GO; GO:0006546; P:glycine catabolism; IEA.
DR InterPro; IPR006223; GCVT.
DR InterPro; IPR006222; GCVT.
DR Pfam; PF01571; GCV_T; 1.
DR TIGRFAMs; TIGR00528; gcvT; 1.
KW Transferase; Complete proteome.
SQ SEQUENCE 379 AA; 40369 MW; 047C7002EF5A30CF CRC64;

Query Match
Best Local Similarity 87.5%; Score 41; DB 16; Length 379;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 5 GGEKAGGF 12
   ||| ||| |||
Db 270 GGERAGGF 277

RESULT 13
Q9YFP0
ID Q9YFP0 PRELIMINARY; PRT; 383 AA.
AC Q9YFP0;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE 383AA long hypothetical S2p metalloprotease.
GN APE0209.
OS Aeropyrum pernix.
OC Archaea; Crenarchaeota; Thermoprotei; Desulfurococcales;
OC Desulfurococaceae; Aeropyrum.
OX NCBI_TaxID=56636;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=X1;
RX MEDLINE=99310339; PubMed=10382966;
RA Kwarababayasi Y., Hino Y., Horikawa H., Yamazaki S., Haikawa Y.,
RA Jin-no K., Takahashi M., Sekine M., Baba S.-I., Ankaï A., Kosugi H.,
RA Hosoyama A., Fukui S., Nagai Y., Nishijima K., Nakazawa H.,
RA Takamiya M., Masuda S., Funahashi T., Tanaka T., Kudoh Y.,
RA Yamazaki J., Kushida N., Oguchi A., Aoki K.-I., Kubota K.,
RA Nakamura Y., Nomura N., Sako Y., Kikuchi H.;
RT "Complete genome sequence of an aerobic hyper-thermophilic
RT crenarchaeon, Aeropyrum pernix K1.";
RL DNA Res. 6:83-101(1999).
DR EMBL; AP000058; BAA79121.1; -.
DR PIR; G72777; G72777.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0004222; F:metalloendopeptidase activity; IEA.
DR GO; GO:0008270; F:zinc ion binding; IEA.
DR GO; GO:0007242; P:intracellular signaling cascade; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR008915; Peptidase M50.
DR InterPro; IPR001193; Pept_M50_SREBP.
DR InterPro; IPR006025; Pept_M_Zn_BS.
DR Pfam; PF00595; PDZ; 1.
DR Pfam; PF02163; Peptidase M50; 1.
DR PRINTS; PR01000; SREPS2PTASE.
DR SMART; SM00228; PDZ; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
KW Protease; Metalloprotease; Complete proteome.
SQ SEQUENCE 383 AA; 39729 MW; F57289CA4F06B8F76 CRC64;
```

Query Match 66.1%; Score 41; DB 17; Length 383;  
Best Local Similarity 80.0%; Pred. No. 1.9e+02;  
Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 GIGGEKAGGF 12  
DB 97 GEGGEAAGGF 106

## RESULT 14

Q8RS19 PRELIMINARY; PRT; 460 AA.  
AC Q8RS19  
DT 01-JUN-2002 (TrEMBLrel. 21, Created)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE TrbI protein.  
GN TrbI.  
OS Gamma-proteobacterium Hot 75m4.  
OG plasmid pB4.  
OC Bacteria; environmental samples.  
OX NCBI\_TaxID=77133;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Tauch A.;  
RT "The Incp beta plasmid pB4 encodes a tripartite antibiotic efflux  
RT system of the RND-MFP-OEP type conferring erythromycin and  
RT roxithromycin resistance in Pseudomonas sp. B13.";  
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AJ431260; CAD24375.1; -  
DR GO; GO:0046821; C:extrachromosomal DNA; IEA.  
DR GO; GO:0009291; P:unidirectional conjugation; IEA.  
DR InterPro; IPR005498; TrbI.  
DR Pfam; PF03743; TrbI; 1.  
KW Plasmid.  
SQ SEQUENCE 460 AA; 47889 MW; B88772672063B91D CRC64;

Query Match 66.1%; Score 41; DB 2; Length 460;  
Best Local Similarity 60.0%; Pred. No. 2.2e+02;  
Matches 6; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 2 AGIGGEKAGG 11  
DB 211 SGVGGQAGG 220

## RESULT 15

Q27376 PRELIMINARY; PRT; 974 AA.  
AC Q27376  
DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE GLH-2.  
GN GLH-2.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=N2;  
RA Grudl M.E.; Smith P.A.; Kuznicki K.A.; McCrone J.S.; Kirchner J.;  
RA Strome S.; Bennett K.L.;  
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.  
CC -!- SIMILARITY: BELONGS TO THE DEAD BOX HELICASE FAMILY.  
DR EMBL; U60449; AAB03510.1; -  
DR EMBL; U60194; AAB03337.1; -  
DR HSSP; Q58083; 1HV8.  
DR GO; GO:0005524; F:ATP binding; IEA.  
DR GO; GO:0008026; F:ATP dependent helicase activity; IEA.  
DR GO; GO:0016787; F:hydrolase activity; IEA.  
DR GO; GO:0003676; F:nucleic acid binding; IEA.

DR InterPro; IPR001410; DEAD  
DR InterPro; IPR000629; DEAD\_box.  
DR InterPro; IPR001650; Helicase\_C.  
DR InterPro; IPR001878; Znf\_CCHC.  
DR Pfam; PF00270; DEAD; 1.  
DR Pfam; PF00271; helicase\_C; 1.  
DR Pfam; PF00098; zf-CCHC; 6.  
DR PRINTS; PR00939; C2HCZNFINGER.  
DR SMART; SM00487; DEAD; 1.  
DR SMART; SM00490; HELIC; 1.  
DR SMART; SM00343; Znf\_C2HC; 6.  
DR PROSITE; PS00039; DEAD ATP HELICASE; 1.  
DR PROSITE; PS0158; ZF\_CCHC; 6.  
KW ATP-binding; Helicase; Hydrolase.  
SQ SEQUENCE 974 AA; 100426 MW; BF994CA586D86C71 CRC64;

Query Match 66.1%; Score 41; DB 5; Length 974;  
Best Local Similarity 63.6%; Pred. No. 4.9e+02;  
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 AGIGGEKAGGF 12  
DB 214 SGFGGKSGGF 224

Search completed: September 18, 2004, 04:29:42  
Job time : 43.5882 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 18, 2004, 03:47:39 ; Search time 26.7647 Seconds  
(without alignments)  
52.784 Million cell updates/sec

Title: US-10-615-959-44  
Perfect score: 26  
Sequence: 1 EKAGG 5

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A. Geneseq\_29Jan04.\*  
1: geneseqp1980s.\*  
2: geneseqp1990s.\*  
3: geneseqp2000s.\*  
4: geneseqp2001s.\*  
5: geneseqp2002s.\*  
6: geneseqp2003as.\*  
7: geneseqp2003bs.\*  
8: geneseqp2004s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	26	100.0	6	2 AAR12514	Aar12514 Part of t
2	26	100.0	6	4 AAB61739	Aab61739 Human typ
3	26	100.0	7	2 AAR71720	Aar71720 Potential
4	26	100.0	7	3 AAY96119	Aay96119 Collagen
5	26	100.0	7	4 AAB61721	Aab61721 Human typ
6	26	100.0	7	5 AAE16472	Aae16472 Human col
7	26	100.0	7	5 ABB80730	Abb80730 Collagen
8	26	100.0	7	5 ABB09622	Abb09622 Collagen
9	26	100.0	8	4 AAB61722	Aab61722 Human typ
10	26	100.0	8	4 AAB61740	Aab61740 Human typ
11	26	100.0	8	4 AAB61724	Aab61724 Human typ
12	26	100.0	9	4 AAB61725	Aab61725 Human typ
13	26	100.0	9	4 AAB61723	Aab61723 Human typ
14	26	100.0	9	4 AAB61741	Aab61741 Human typ
15	26	100.0	10	4 AAB61726	Aab61726 Human typ
16	26	100.0	10	7 ADA37452	Ada37452 Human typ
17	26	100.0	11	4 AAB61731	Aab61731 Human typ
18	26	100.0	11	4 AAB61742	Aab61742 Human typ
19	26	100.0	12	4 AAB61730	Aab61730 Human typ
20	26	100.0	12	4 AAB61743	Aab61743 Human typ
21	26	100.0	13	4 AAB61727	Aab61727 Human typ
22	26	100.0	14	2 AAW98115	Aaw98115 T-cell me
23	26	100.0	16	4 AAB61728	Aab61728 Human typ
24	26	100.0	16	4 ABB56493	Abb56493 Human MHC
25	26	100.0	17	4 AAB61729	Aab61729 Human typ

26	26	100.0	17	4 ABB56496	Abb56496 Human MHC
27	26	100.0	19	7 AAR38626	Aae38626 Human N-t
28	26	100.0	21	4 AAM98908	Aam98908 Vaccine r
29	26	100.0	24	6 ABR43102	Abr43102 Human T-c
30	26	100.0	24	6 AAE37759	Aae37759 TIRC7 pep
31	26	100.0	24	6 AAE37809	Aae37809 TIRC7 pep
32	26	100.0	25	7 AAE38632	Aae38632 Human C-t
33	26	100.0	30	2 AAR92068	Aar92068 Recombina
34	26	100.0	34	2 AAR92065	Aar92065 Recombina
35	26	100.0	51	5 ABP33968	Abp33968 Human ORF
36	26	100.0	59	3 AAG15968	Aag15968 Arabidops
37	26	100.0	65	4 AAM19314	Aam19314 Peptide #
38	26	100.0	65	4 AAB38630	Abb38630 Peptide #
39	26	100.0	65	4 AAM32088	Aam32088 Peptide #
40	26	100.0	65	4 ABE23723	Abb23723 Protein #
41	26	100.0	65	4 AAM71802	Aam71802 Human bon
42	26	100.0	65	4 AAM59260	Aam59260 Human bra
43	26	100.0	65	4 ABG53486	Abg53486 Human liv
44	26	100.0	65	5 ABG41615	Abg41615 Human pep
45	26	100.0	68	3 AAB24662	Aab24662 Plant SDF

ALIGNMENTS

RESULT 1  
AAR12514  
ID AAR12514 standard; peptide; 6 AA.  
XX AC AAR12514;  
XX  
XX 25-MAR-2003 (revised)  
DT 12-SEP-1991 (first entry)  
XX  
XX Part of type III collagen telopeptide.  
XX Degradation; inflammation; cross link; osteoarthritis.  
XX Homo sapiens.  
XX  
XX Key Location/Qualifiers  
FH Modified-site 2  
FT /note= "see comments"  
ET  
XX  
XX MO9108478-A.  
XX  
XX 13-JUN-1991.  
XX  
XX 01-DEC-1989; 89US-00444881.  
XX  
XX 01-DEC-1989; 89US-00444881.  
XX 21-NOV-1990; 90US-00614719.  
XX (WASH-) WASHINGTON RES FOUND.  
XX  
XX Eyre DR;  
XX WPI; 1991-193334/26.  
XX  
XX Detection of collagen degradation peptide(s) - used for detecting, diagnosing and monitoring degenerative and inflammatory disorders.  
XX Claim 12; Page 39; 58pp; English.  
XX  
XX The peptide is part of a larger molecule to which it is linked via the lysine in the 2 posn: Glu-Hyl-Ala-Gly-Phe... (AAR12514) CC | Glu-Hyl-Ala-Gly-Phe CC | Hyl Hyl-Hyl-Hyl = hydroxylysyl pyridinolone. The peptide, or a similar molecule having the pyridinium ring cleaved, can be detected in body fluid using a specific binding partner. The amt. present gives an indication of bone resorption and can be used to screen for osteoarthritis and similar degenerative and inflammatory disorders. The peptide was prepd. from human urine using reverse phase and molecular sieve chromatography. See also AAR12510-R12513. (Updated on 25-MAR-2003

CC to correct PA field.)  
XX  
SQ Sequence 6 AA;  
Query Match 100.0%; Score 26; DB 2; Length 6;  
Best Local Similarity 100.0%; Pred. No. 1.4e+06; Indels 0;  
Matches 5; Conservative 0; Mismatches 0; Gaps 0;  
QY 1 EKAGG 5  
DB 1 EKAGG 5  
RESULT 2  
AAB61739  
ID AAB61739 standard; peptide; 6 AA.  
XX  
AC AAB61739;  
XX  
DT 20-APR-2001 (first entry)  
XX  
DE Human type III collagen carboxy-telopeptide fragment.  
XX  
KW Type II collagen; immunoassay; cartilage; telopeptide; human;  
XX type III collagen.  
XX  
OS Synthetic.  
OS Homo sapiens.  
XX  
FN WO200079284-A1.  
XX  
PD 28-DEC-2000.  
XX  
PF 10-DEC-1999; 99WO-US029357.  
XX  
PR 17-JUN-1999; 99US-00335098.  
PR 29-JUN-1999; 99US-0141574P.  
PR 02-JUL-1999; 99US-0142274P.  
PR 07-JUL-1999; 99US-0142675P.  
PR 30-AUG-1999; 99US-00385740.  
XX  
PA (WASH-) WASHINGTON RES FOUND.  
XX  
PI Eyre DR;  
XX  
DR WPI; 2001-146859/15.  
XX  
PT Assay for detecting cross-linked telopeptide analytes indicative of type II collagen resorption in vivo in a body fluid sample, comprises contacting the sample with an antibody which binds to the analyte.  
XX  
PS Disclosure; Page 15; 34pp; English.  
XX  
CC The invention relates to immunoassays for measuring type II collagen (cartilage) resorption in vivo. The method of analysing a body fluid sample for the presence of an analyte indicative of a physiological condition, involves contacting the body fluid sample with an antibody (Ab) which binds to the analyte, detecting binding of Ab in the body fluid sample, and correlating any detected binding to the physiological condition. The analysis is useful for measuring type II collagen (cartilage) resorption in vivo, for distinguishing between resorption of non-mineralized and mineralized cartilage, and for measuring total cartilage resorption in vivo. The present sequence represents a linear synthetic peptide based on the carboxy-terminal telopeptide sequence of human collagen type III  
XX  
SQ Sequence 6 AA;  
Query Match 100.0%; Score 26; DB 4; Length 6;  
Best Local Similarity 100.0%; Pred. No. 1.4e+06;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 EKAGG 5  
DB 1 EKAGG 5  
RESULT 3  
AAR71720  
ID AAR71720 standard; protein; 7 AA.  
XX  
AC AAR71720;  
XX  
DT 25-MAR-2003 (revised)  
DT 18-OCT-1995 (first entry)  
XX  
DE Potential cross-linking sequence of collagen type III.  
XX  
KW Collagen; antibody; immunoassay; metabolism; diagnosis; monitoring; disorder; osteoporosis; metastatic progression; Paget's disease; hyperthyroidism; bone; resorption; rheumatoid arthritis; osteoarthritis; vasculitis syndrome.  
XX  
OS Homo sapiens.  
XX  
FN WO9508115-A1.  
XX  
PD 23-MAR-1995.  
XX  
PF 19-SEP-1994; 94WO-DK000348.  
XX  
PR 17-SEP-1993; 93DK-00001040.  
XX  
PA (OSTE-) OSTEOMETER AS.  
XX  
PI Qvist P, Bonde M;  
XX  
DR WPI; 1995-131456/17.  
XX  
PT Assaying collagen fragments in body fluid by immunoassay - using antibodies raised against synthetic peptide(s) contg. potential crosslinking sites, to diagnose and monitor disorders of collagen metabolism, e.g. osteoporosis.  
XX  
PS Claim 7; Page 57; 87pp; English.  
XX  
CC Determination of collagen fragments in body fluids can be achieved by immunoassay using antibodies directed against synthetic peptides derived from collagen which contain sites of potential crosslinking. The method is used to diagnose and monitor treatment of disorders of collagen metabolism (degradation of type I collagen may indicate osteoporosis, metastatic progression, Paget's disease, hyperthyroidism or other conditions involving excessive bone resorption; degradation of type II collagen may indicate rheumatoid arthritis or osteoarthritis; and of type III collagen, vasculitis syndrome). The method can also be used to assess the toxicity of a compound and to test drugs for their effect on collagen metabolism. The sequences of collagen fragments on which the synthetic peptides were based are described in AAR71705-07 and AAR71709-22. This telopeptide sequence is a potential cross-linking site in collagen alpha1 (type III) and is found at the C-terminus. (Updated on 25-MAR-2003 to correct FN field.)  
XX  
SQ Sequence 7 AA;  
Query Match 100.0%; Score 26; DB 2; Length 7;  
Best Local Similarity 100.0%; Pred. No. 1.4e+06;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 EKAGG 5  
DB 1 EKAGG 5  
RESULT 4  
AAY96119  
ID AAY96119 standard; peptide; 7 AA.



XX AC AAY96119;  
XX DT 19-DEC-2000 (first entry)  
XX XX  
DE Collagen type III alpha-1 C-terminal peptide.  
XX XX  
KW Collagen type III; vasculitis syndrome; assay; diagnosis.  
XX OS Homo sapiens.  
XX PN US6110689-A.  
XX PD 29-AUG-2000.  
XX PF 04-NOV-1997; 97US-00963825.  
XX PR 21-JAN-1994; 94US-00187319.  
XX PA (OSTE-) OSTEOMETER AS.  
XX PI Bonde M, Qvist P;  
XX DR WPI; 2000-586349/55.  
XX XX  
PT Assaying type I collagen fragments for diagnosing osteoporosis in  
PT postmenopausal woman, involves contacting body fluid with synthetic  
PT collagen peptide and antibody and quantifying by competitive binding  
PT assay.  
XX PS Disclosure; Col 9; 41pp; English.  
XX XX  
CC This synthetic peptide is derived from the C-terminus of type III  
CC collagen alpha-1 (see AAY96125). It represents a potential crosslinkable  
CC site of collagen III. The invention is based on the discovery of the  
CC presence of particular collagen fragments in body fluids of patients  
CC compared with those of healthy subjects. These fragments are generated  
CC upon collagen degradation and are partly characterised by the presence of  
CC potential sites for crosslinking. A method for assaying collagen  
CC fragments in a body fluid sample is based on the competitive binding of  
CC collagen fragments in the sample and of synthetic peptides derived from  
CC collagen to immunological binding partners. When considering the  
CC degradation of type III collagen, the assay can be used as a means of  
CC identifying the presence of vasculitis syndrome. The assay can also be  
CC used to monitor the progress of therapy  
XX SQ  
SQ Sequence 7 AA;  
Query Match 100.0%; Score 26; DB 3; Length 7;  
Best Local Similarity 100.0%; Pred. NO. 1.4e+06;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 EKAGG 5  
Db 1 EKAGG 5  
RESULT 5  
AAB61721  
ID AAB61721 standard; peptide; 7 AA.  
XX XX  
AC AAB61721;  
XX DT 20-APR-2001 (first entry)  
XX DE Human type II collagen carboxy-telopeptide fragment.  
XX XX  
KW Type II collagen; immunoassay; cartilage; telopeptide; human.  
XX OS Synthetic.  
XX OS Homo sapiens.  
XX XX  
PN W0200079284-A1.

XX ED 28-DEC-2000.  
XX PF 10-DEC-1999; 99WO-US029357.  
XX XX  
PR 17-JUN-1999; 99US-00335098.  
PR 29-JUN-1999; 99US-0141574P.  
PR 02-JUL-1999; 99US-0142274P.  
PR 07-JUL-1999; 99US-0142675P.  
PR 30-AUG-1999; 99US-00385740.  
XX XX  
PA (WASH-) WASHINGTON RES FOUND.  
XX XX  
PI Eyre DR;  
XX DR WPI; 2001-146859/15.  
XX XX  
PT Assay for detecting cross-linked telopeptide analytes indicative of type  
PT II collagen resorption in vivo in a body fluid sample, comprises  
PT contacting the sample with an antibody which binds to the analyte.  
XX XX  
PS Claim 1; Page 16; 34pp; English.  
XX XX  
CC The invention relates to immunoassays for measuring type II collagen  
CC (cartilage) resorption in vivo. The method of analysing a body fluid  
CC sample for the presence of an analyte indicative of a physiological  
CC condition, involves contacting the body fluid sample with an antibody  
CC (Ab) which binds to the analyte, detecting binding of Ab in the body  
CC fluid sample, and correlating any detected binding to the physiological  
CC condition. The analysis is useful for measuring type II collagen  
CC (cartilage) resorption in vivo, for distinguishing between resorption of  
CC non-mineralized and mineralized cartilage, and for measuring total  
CC cartilage resorption in vivo. The present sequence represents a peptide  
CC based on the carboxy-terminal telopeptide sequence of human collagen type  
CC II, to which an antibody binds to  
XX SQ  
SQ Sequence 7 AA;  
Query Match 100.0%; Score 26; DB 4; Length 7;  
Best Local Similarity 100.0%; Pred. NO. 1.4e+06;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 EKAGG 5  
Db 2 EKAGG 6  
RESULT 6  
AAB616472  
ID AAB616472 standard; peptide; 7 AA.  
XX XX  
AC AAB616472;  
XX DT 09-APR-2002 (first entry)  
XX DE Human collagen type III alpha (III) C-terminal peptide.  
XX XX  
KW Human; collagen type III; osteoporosis; bone resorption; Paget's disease;  
KW hyperparathyroidism; rheumatoid arthritis; osteoarthritis; therapy;  
KW vasculitis syndrome; toxic substance; tissue degradation; alpha(III).  
XX OS Homo sapiens.  
XX XX  
PN US6323314-B1.  
XX PD 27-NOV-2001.  
XX XX  
PF 10-FEB-2000; 2000US-00500811.  
XX PR 21-JAN-1994; 94US-00187319.  
XX PR 04-NOV-1997; 97US-00963825.  
XX XX  
PA (OSTE-) OSTEOMETER AS.

```

XX Qvist P, Bonde M;
XX WPI; 2002-096598/13.
XX Novel peptides that match alpha 1 or 2(I) telopeptide component of cross-
XX linked telopeptide degradation product of type I collagen useful for
XX diagnosing disorders associated with collagen metabolism e.g.
XX osteoporosis.
XX Example 1; Col 8; 34pp; English.
XX The invention relates to a peptide synthesised to match an alpha(I) or
XX alpha2(I) telopeptide component of a cross-linked telopeptide degradation
XX product of type I collagen. The method is useful for assaying collagen
XX fragments in animal body fluids, for determining the degradation of human
XX collagen types I, II and III and for diagnosing the presence of disorders
XX associated with the metabolism of collagen, especially osteoporosis. The
XX peptide is used for assessing the impact of drugs on collagen metabolism.
XX The peptide is useful in methods to assess an abnormal condition of a
XX subject for e.g. excessive bone resorption which shows the presence of a
XX osteoporotic condition or the metastatic progress of a malignancy,
XX Paget's disease and hyperparathyroidism. Disease states involving
XX connective tissues can be monitored by determining collagen degradation,
XX examples are collagen type II degradation associated with rheumatoid
XX arthritis, osteoarthritis, and collagen type III degradation in
XX vasculitis syndrome. Since the conditions of the subject can be monitored
XX continuously, application of these assays can also be used to monitor the
XX progress of therapy administered to treat these or other conditions and
XX as a measure of toxicity, since the administration of toxic substances
XX often results in tissue degradation. The present sequence is human
XX collagen type III alpha(III) C-terminal peptide used for synthesising a
XX peptide used in the invention
XX Sequence 7 AA;
SQ
Query Match 100.0%; Score 26; DB 5; Length 7;
Best Local Similarity 100.0%; Pred. No. 1.4e+06;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 EXAGG 5
DB |||||
1 EXAGG 5
RESULT 7
ABB80730
ID ABB80730 standard; peptide; 7 AA.
XX ABB80730;
XX 15-JUL-2002 (first entry)
XX Collagen type III-alpha C-terminal fragment.
XX Collagen; osteoarthritis; Paget's disease; Marfan syndrome; dwarfism;
XX osteogenesis imperfecta; neoplastic growth; rheumatoid arthritis;
XX vasculitis.
XX Synthetic.
XX Homo sapiens.
XX US6355442-B1.
XX 12-MAR-2002.
XX 13-APR-2000; 2000US-00548608.
XX 21-JAN-1994; 94US-00187319.
XX 04-NOV-1997; 97US-00963825.
XX (OSTE-) OSTEOMETER BIOTECH AS.
XX
XX Qvist P, Bonde M;
XX WPI; 2002-224940/28.
XX Assaying type II collagen fragments in a body fluid sample, for
XX determining abnormalities in collagen metabolism, e.g. in Paget's disease
XX or Marfan's syndrome, comprises contacting the fluid with a synthetic
XX type II collagen alpha.
XX Disclosure; Col 9-10; 35pp; English.
XX
PI Qvist P, Bonde M;
XX WPI; 2002-380937/41.
XX Assaying type I collagen fragments in body fluid, useful for diagnosis
XX PT and assessing treatment of e.g. osteoarthritis, by competitive
XX PT immunoassay.
XX Disclosure; Col 7-8; 35pp; English.
XX
XX The invention relates to a method for assaying type I collagen fragments
XX (I) in body fluid. The method involves treating the test sample with: (i)
XX synthetic peptide, immobilised on a support; and (ii) immunological
XX binding partner, reactive with the synthetic peptide, so that (i) and the
XX synthetic peptide compete for binding, and (i) are quantified by
XX measuring the binding of the binding partner to the synthetic peptide.
XX The method is used to diagnose disorders of collagen metabolism,
XX especially osteoarthritis but also Paget's disease, Marfan syndrome,
XX osteogenesis imperfecta, neoplastic growth of collagenous tissue,
XX dwarfism, rheumatoid arthritis or vasculitis, also for clinical testing
XX of drugs to assess their effect on collagen metabolism. Sequences
XX ABB80716-732 represent synthetic peptide derived from collagen used for
XX assaying type I, II and III collagen fragments
XX Sequence 7 AA;
SQ
Query Match 100.0%; Score 26; DB 5; Length 7;
Best Local Similarity 100.0%; Pred. No. 1.4e+06;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 EXAGG 5
DB |||||
1 EXAGG 5
RESULT 8
ABB09622
ID ABB09622 standard; peptide; 7 AA.
XX ABB09622;
XX 29-MAY-2002 (first entry)
XX Collagen type III peptide with sites for crosslinking to collagens.
XX Collagen; bone resorption; collagen metabolism; Paget's disease;
XX Marfan's syndrome; osteogenesis imperfecta; neoplastic growth; dwarfism;
XX rheumatoid arthritis; osteoarthritis; vasculitis syndrome.
XX Homo sapiens.
XX US6342361-B1.
XX 29-JAN-2002.
XX 12-MAY-2000; 2000US-00570573.
XX 21-JAN-1994; 94US-00187319.
XX 04-NOV-1997; 97US-00963825.
XX (OSTE-) OSTEOMETER BIOTECH AS.
XX
XX Qvist P, Bonde M;
XX WPI; 2002-224940/28.
XX Assaying type II collagen fragments in a body fluid sample, for
XX determining abnormalities in collagen metabolism, e.g. in Paget's disease
XX or Marfan's syndrome, comprises contacting the fluid with a synthetic
XX type II collagen alpha.
XX Disclosure; Col 9-10; 35pp; English.
XX

```

CC The present sequence represents a peptide derived from collagen type III  
 CC alpha1. The peptide has potential sites for crosslinking to various  
 CC collagens. The specification describes a method for assaying type II  
 CC collagen fragments in a body fluid sample. The method comprises  
 CC contacting the body fluid with a synthetic peptide consisting essentially  
 CC of a type II collagen alpha1 amino acid sequence that competes with the  
 CC collagen fragments for binding with the immunological binding partner.  
 CC The method is useful for assaying collagen fragments in body fluids. The  
 CC method is particularly useful in assays for measuring bone resorption  
 CC rates in patients or in healthy subjects. The method may be used for  
 CC determining abnormalities in collagen metabolism, especially in Paget's  
 CC disease, Marfan's syndrome, osteogenesis imperfecta, neoplastic growth in  
 CC collagenous tissue, dwarfism, rheumatoid arthritis, osteoarthritis or  
 CC vasculitis syndrome. The method may be used for determining the  
 CC degradation of human collagen of type I, II and III. It also can be used  
 CC during clinical testing of new drugs to assess the impact of these drugs  
 CC on collagen metabolism. The assays can be used as a measure of toxicity,  
 CC since the administration of toxic substances often results in tissue  
 CC degradation. Thus, the assays may be applied in any situation where the  
 CC metabolic condition of collagen tissues can be used as an index of the  
 CC condition, treatment or effect of substances directly administered to the  
 CC subject or to which the subject is exposed in the environment  
 XX  
 XX SQ Sequence 7 AA;

Query Match 100.0%; Score 26; DB 5; Length 7;  
 Best Local Similarity 100.0%; Pred. No. 1.4e+06;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 EKAGG 5  
 |||||  
 Db 1 EKAGG 5

RESULT 9  
 AAB61722  
 ID AAB61722 standard; peptide; 8 AA.  
 XX  
 AC AAB61722;  
 XX  
 DT 20-APR-2001 (first entry)  
 XX  
 DE Human type II collagen carboxy-telopeptide fragment.  
 XX  
 KW Type II collagen; immunoassay; cartilage; telopeptide; human.

XX  
 OS Synthetic.  
 OS Homo sapiens.  
 XX  
 PN WO200079284-A1.  
 XX  
 PD 28-DEC-2000.  
 XX  
 PF 10-DEC-1999; 99WO-US029357.  
 XX  
 PR 17-JUN-1999; 99US-00335098.  
 PR 29-JUN-1999; 99US-0141574P.  
 PR 02-JUL-1999; 99US-0142274P.  
 PR 07-JUL-1999; 99US-0142675P.  
 PR 30-AUG-1999; 99US-00385740.  
 XX  
 PA (WASH-) WASHINGTON RES FOUND.  
 XX  
 PI Eyre DR;  
 XX  
 DR WPI; 2001-146859/15.

XX Assay for detecting cross-linked telopeptide analytes indicative of type  
 XX II collagen resorption in vivo in a body fluid sample, comprises  
 XX contacting the sample with an antibody which binds to the analyte.  
 XX  
 XX Disclosure; Page 7; 34pp; English.

CC The invention relates to immunoassays for measuring type II collagen  
 CC (cartilage) resorption in vivo. The method of analysing a body fluid  
 CC sample for the presence of an analyte indicative of a physiological  
 CC condition, involves contacting the body fluid sample with an antibody  
 CC (Ab) which binds to the analyte, detecting binding of Ab in the body  
 CC fluid sample, and correlating any detected binding to the physiological  
 CC condition. The analysis is useful for measuring type II collagen  
 CC (cartilage) resorption in vivo, for distinguishing between resorption of  
 CC non-mineralized and mineralized cartilage, and for measuring total  
 CC cartilage resorption in vivo. The present sequence represents a linear  
 CC synthetic peptide based on the carboxy-terminal telopeptide sequence of  
 CC human collagen type II  
 XX  
 XX SQ Sequence 8 AA;

Query Match 100.0%; Score 26; DB 4; Length 8;  
 Best Local Similarity 100.0%; Pred. No. 1.4e+06;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 EKAGG 5  
 |||||  
 Db 2 EKAGG 6

RESULT 10  
 AAB61740  
 ID AAB61740 standard; peptide; 8 AA.  
 XX  
 AC AAB61740;

XX  
 DT 20-APR-2001 (first entry)  
 XX  
 DE Human type III collagen carboxy-telopeptide fragment.  
 XX  
 KW Type II collagen; immunoassay; cartilage; telopeptide; human;  
 XX type III collagen.

XX  
 OS Synthetic.  
 OS Homo sapiens.  
 XX  
 PN WO200079284-A1.  
 XX  
 PD 28-DEC-2000.

XX  
 PF 10-DEC-1999; 99WO-US029357.  
 XX  
 PR 17-JUN-1999; 99US-00335098.  
 PR 29-JUN-1999; 99US-0141574P.  
 PR 02-JUL-1999; 99US-0142274P.  
 PR 07-JUL-1999; 99US-0142675P.  
 PR 30-AUG-1999; 99US-00385740.  
 XX  
 PA (WASH-) WASHINGTON RES FOUND.

XX  
 PI Eyre DR;  
 XX  
 DR WPI; 2001-146859/15.

XX Assay for detecting cross-linked telopeptide analytes indicative of type  
 XX II collagen resorption in vivo in a body fluid sample, comprises  
 XX contacting the sample with an antibody which binds to the analyte.

XX  
 PS Disclosure; Page 15; 34pp; English.

XX The invention relates to immunoassays for measuring type II collagen  
 CC (cartilage) resorption in vivo. The method of analysing a body fluid  
 CC sample for the presence of an analyte indicative of a physiological  
 CC condition, involves contacting the body fluid sample with an antibody  
 CC (Ab) which binds to the analyte, detecting binding of Ab in the body  
 CC fluid sample, and correlating any detected binding to the physiological  
 CC condition. The analysis is useful for measuring type II collagen  
 CC (cartilage) resorption in vivo, for distinguishing between resorption of  
 CC non-mineralized and mineralized cartilage, and for measuring total

CC cartilage resorption in vivo. The present sequence represents a linear  
 CC synthetic peptide based on the carboxy-terminal telopeptide sequence of  
 CC human collagen type III  
 XX  
 SQ Sequence 8 AA;  
 Query Match 100.0%; Score 26; DB 4; Length 8;  
 Best Local Similarity 100.0%; Pred. No. 1.4e+06;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 EKAGG 5  
 Db |||||  
 4 EKAGG 8  
 RESULT 11  
 AAB61724  
 ID AAB61724 standard; peptide; 8 AA.  
 XX AAB61724;  
 AC AAB61724;  
 XX 20-APR-2001 (first entry)  
 DT Human type II collagen carboxy-telopeptide fragment.  
 XX Type II collagen; immunoassay; cartilage; telopeptide; human.  
 DE Synthetic.  
 KW Homo sapiens.  
 OS WO200079284-A1.  
 XX 28-DEC-2000.  
 PD 10-DEC-1999; 99WO-US029357.  
 PF 17-JUN-1999; 99US-00335098.  
 XX 29-JUN-1999; 99US-0141574P.  
 PR 02-JUL-1999; 99US-0142274P.  
 PR 07-JUL-1999; 99US-0142675P.  
 PR 30-AUG-1999; 99US-00385740.  
 XX (WASH-) WASHINGTON RES FOUND.  
 PA Eyre DR;  
 PI WPI; 2001-146859/15.  
 XX Assay for detecting cross-linked telopeptide analytes indicative of type  
 PT II collagen resorption in vivo in a body fluid sample, comprises  
 PT contacting the sample with an antibody which binds to the analyte.  
 XX Disclosure; Page 7; 34pp; English.  
 PS The invention relates to immunoassays for measuring type II collagen  
 CC (cartilage) resorption in vivo. The method of analysing a body fluid  
 CC sample for the presence of an analyte indicative of a physiological  
 CC condition, involves contacting the body fluid sample with an antibody  
 CC (Ab) which binds to the analyte, detecting binding of Ab in the body  
 CC fluid sample, and correlating any detected binding to the physiological  
 CC condition. The analysis is useful for measuring type II collagen  
 CC (cartilage) resorption in vivo, for distinguishing between resorption of  
 CC non-mineralized and mineralized cartilage, and for measuring total  
 CC cartilage resorption in vivo. The present sequence represents a linear  
 CC synthetic peptide based on the carboxy-terminal telopeptide sequence of  
 CC human collagen type II  
 XX Sequence 8 AA;  
 Query Match 100.0%; Score 26; DB 4; Length 8;  
 Best Local Similarity 100.0%; Pred. No. 1.4e+06;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 EKAGG 5  
 Db |||||  
 4 EKAGG 8  
 RESULT 12  
 AAB61725  
 ID AAB61725 standard; peptide; 9 AA.  
 XX AAB61725;  
 AC AAB61725;  
 XX 20-APR-2001 (first entry)  
 DT Human type II collagen carboxy-telopeptide fragment.  
 XX Type II collagen; immunoassay; cartilage; telopeptide; human.  
 DE Synthetic.  
 KW Homo sapiens.  
 OS WO200079284-A1.  
 XX 28-DEC-2000.  
 PD 10-DEC-1999; 99WO-US029357.  
 PF 17-JUN-1999; 99US-00335098.  
 XX 29-JUN-1999; 99US-0141574P.  
 PR 02-JUL-1999; 99US-0142274P.  
 PR 07-JUL-1999; 99US-0142675P.  
 PR 30-AUG-1999; 99US-00385740.  
 XX (WASH-) WASHINGTON RES FOUND.  
 PA Eyre DR;  
 PI WPI; 2001-146859/15.  
 XX Assay for detecting cross-linked telopeptide analytes indicative of type  
 PT II collagen resorption in vivo in a body fluid sample, comprises  
 PT contacting the sample with an antibody which binds to the analyte.  
 XX Disclosure; Page 7; 34pp; English.  
 PS The invention relates to immunoassays for measuring type II collagen  
 CC (cartilage) resorption in vivo. The method of analysing a body fluid  
 CC sample for the presence of an analyte indicative of a physiological  
 CC condition, involves contacting the body fluid sample with an antibody  
 CC (Ab) which binds to the analyte, detecting binding of Ab in the body  
 CC fluid sample, and correlating any detected binding to the physiological  
 CC condition. The analysis is useful for measuring type II collagen  
 CC (cartilage) resorption in vivo, for distinguishing between resorption of  
 CC non-mineralized and mineralized cartilage, and for measuring total  
 CC cartilage resorption in vivo. The present sequence represents a linear  
 CC synthetic peptide based on the carboxy-terminal telopeptide sequence of  
 CC human collagen type II  
 XX Sequence 9 AA;  
 Query Match 100.0%; Score 26; DB 4; Length 9;  
 Best Local Similarity 100.0%; Pred. No. 1.4e+06;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 EKAGG 5  
 Db |||||  
 3 EKAGG 7  
 RESULT 13  
 AAB61723  
 ID AAB61723 standard; peptide; 9 AA.  
 XX AAB61723;  
 AC AAB61723;

QY 1 EKAGG 5  
 Db |||||  
 3 EKAGG 7  
 RESULT 12  
 AAB61725  
 ID AAB61725 standard; peptide; 9 AA.  
 XX AAB61725;  
 AC AAB61725;  
 XX 20-APR-2001 (first entry)  
 DT Human type II collagen carboxy-telopeptide fragment.  
 XX Type II collagen; immunoassay; cartilage; telopeptide; human.  
 DE Synthetic.  
 KW Homo sapiens.  
 OS WO200079284-A1.  
 XX 28-DEC-2000.  
 PD 10-DEC-1999; 99WO-US029357.  
 PF 17-JUN-1999; 99US-00335098.  
 XX 29-JUN-1999; 99US-0141574P.  
 PR 02-JUL-1999; 99US-0142274P.  
 PR 07-JUL-1999; 99US-0142675P.  
 PR 30-AUG-1999; 99US-00385740.  
 XX (WASH-) WASHINGTON RES FOUND.  
 PA Eyre DR;  
 PI WPI; 2001-146859/15.  
 XX Assay for detecting cross-linked telopeptide analytes indicative of type  
 PT II collagen resorption in vivo in a body fluid sample, comprises  
 PT contacting the sample with an antibody which binds to the analyte.  
 XX Disclosure; Page 7; 34pp; English.  
 PS The invention relates to immunoassays for measuring type II collagen  
 CC (cartilage) resorption in vivo. The method of analysing a body fluid  
 CC sample for the presence of an analyte indicative of a physiological  
 CC condition, involves contacting the body fluid sample with an antibody  
 CC (Ab) which binds to the analyte, detecting binding of Ab in the body  
 CC fluid sample, and correlating any detected binding to the physiological  
 CC condition. The analysis is useful for measuring type II collagen  
 CC (cartilage) resorption in vivo, for distinguishing between resorption of  
 CC non-mineralized and mineralized cartilage, and for measuring total  
 CC cartilage resorption in vivo. The present sequence represents a linear  
 CC synthetic peptide based on the carboxy-terminal telopeptide sequence of  
 CC human collagen type II  
 XX Sequence 9 AA;  
 Query Match 100.0%; Score 26; DB 4; Length 9;  
 Best Local Similarity 100.0%; Pred. No. 1.4e+06;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 EKAGG 5  
 Db |||||  
 3 EKAGG 7  
 RESULT 13  
 AAB61723  
 ID AAB61723 standard; peptide; 9 AA.  
 XX AAB61723;  
 AC AAB61723;

```

XX 20-APR-2001 (first entry)
XX Human type II collagen carboxy-telopeptide fragment.
XX Type II collagen; immunoassay; cartilage; telopeptide; human.
XX Synthetic.
XX Homo sapiens.
XX WO200079284-A1.
XX 28-DEC-2000.
XX 10-DEC-1999; 99WO-US029357.
XX 17-JUN-1999; 99US-00335098.
XX 29-JUN-1999; 99US-0141574P.
XX 02-JUL-1999; 99US-0142274P.
XX 07-JUL-1999; 99US-0142675P.
XX 30-AUG-1999; 99US-00385740.
XX (WASH-) WASHINGTON RES FOUND.
XX Eyre DR;
XX WPI; 2001-146859/15.
XX Assay for detecting cross-linked telopeptide analytes indicative of type
XX II collagen resorption in vivo in a body fluid sample, comprises
XX contacting the sample with an antibody which binds to the analyte.
XX Disclosure; Page 7; 34pp; English.
XX The invention relates to immunoassays for measuring type II collagen
XX (cartilage) resorption in vivo. The method of analysing a body fluid
XX sample for the presence of an analyte indicative of a physiological
XX condition, involves contacting the body fluid sample with an antibody
XX (Ab) which binds to the analyte, detecting binding of Ab in the body
XX fluid sample, and correlating any detected binding to the physiological
XX condition. The analysis is useful for measuring type II collagen
XX (cartilage) resorption in vivo, for distinguishing between resorption of
XX non-mineralized and mineralized cartilage, and for measuring total
XX cartilage resorption in vivo. The present sequence represents a linear
XX synthetic peptide based on the carboxy-terminal telopeptide sequence of
XX human collagen type II
XX Sequence 9 AA;
XX Query Match 100.0%; Score 26; DB 4; Length 9;
XX Best Local Similarity 100.0%; Pred. No. 1.4e+06;
XX Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX Qy 1 EKAGG 5
XX Db 2 EKAGG 6
XX RESULT 14
XX AAB61741
XX ID AAB61741 standard; peptide; 9 AA.
XX AC AAB61741;
XX XX 20-APR-2001 (first entry)
XX Human type III collagen carboxy-telopeptide fragment.
XX Type II collagen; immunoassay; cartilage; telopeptide; human;
XX Type III collagen.
XX Synthetic.
XX Homo sapiens.

```

```

XX WO200079284-A1.
XX 28-DEC-2000.
XX 10-DEC-1999; 99WO-US029357.
XX 17-JUN-1999; 99US-00335098.
XX 29-JUN-1999; 99US-0141574P.
XX 02-JUL-1999; 99US-0142274P.
XX 07-JUL-1999; 99US-0142675P.
XX 30-AUG-1999; 99US-00385740.
XX (WASH-) WASHINGTON RES FOUND.
XX Eyre DR;
XX WPI; 2001-146859/15.
XX Assay for detecting cross-linked telopeptide analytes indicative of type
XX II collagen resorption in vivo in a body fluid sample, comprises
XX contacting the sample with an antibody which binds to the analyte.
XX Disclosure; Page 15; 34pp; English.
XX The invention relates to immunoassays for measuring type II collagen
XX (cartilage) resorption in vivo. The method of analysing a body fluid
XX sample for the presence of an analyte indicative of a physiological
XX condition, involves contacting the body fluid sample with an antibody
XX (Ab) which binds to the analyte, detecting binding of Ab in the body
XX fluid sample, and correlating any detected binding to the physiological
XX condition. The analysis is useful for measuring type II collagen
XX (cartilage) resorption in vivo, for distinguishing between resorption of
XX non-mineralized and mineralized cartilage, and for measuring total
XX cartilage resorption in vivo. The present sequence represents a linear
XX synthetic peptide based on the carboxy-terminal telopeptide sequence of
XX human collagen type III
XX Sequence 9 AA;
XX Query Match 100.0%; Score 26; DB 4; Length 9;
XX Best Local Similarity 100.0%; Pred. No. 1.4e+06;
XX Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX Qy 1 EKAGG 5
XX Db 4 EKAGG 8
XX RESULT 15
XX AAB61726
XX ID AAB61726 standard; peptide; 10 AA.
XX AC AAB61726;
XX XX 20-APR-2001 (first entry)
XX Human type II collagen carboxy-telopeptide fragment.
XX Type II collagen; immunoassay; cartilage; telopeptide; human.
XX Synthetic.
XX Homo sapiens.
XX WO200079284-A1.
XX 28-DEC-2000.
XX 10-DEC-1999; 99WO-US029357.
XX 17-JUN-1999; 99US-00335098.
XX 29-JUN-1999; 99US-0141574P.
XX 02-JUL-1999; 99US-0142274P.

```

PR 07-JUL-1999; 99US-0142675P.  
PR 30-AUG-1999; 99US-00385740.  
XX  
XX  
XX (WASH-) WASHINGTON RES FOUND.  
XX PI  
XX PI  
XX PI  
XX PI  
XX PI  
XX WPI; 2001-146859/15.  
XX  
XX Assay for detecting cross-linked telopeptide analytes indicative of type  
PT II collagen resorption in vivo in a body fluid sample, comprises  
PT contacting the sample with an antibody which binds to the analyte.  
XX  
PS Disclosure; Page 7; 34pp; English.  
XX  
XX The invention relates to immunoassays for measuring type II collagen  
CC (cartilage) resorption in vivo. The method of analysing a body fluid  
CC sample for the presence of an analyte indicative of a physiological  
CC condition, involves contacting the body fluid sample with an antibody  
CC (Ab) which binds to the analyte, detecting binding of Ab in the body  
CC fluid sample, and correlating any detected binding to the physiological  
CC condition. The analysis is useful for measuring type II collagen  
CC (cartilage) resorption in vivo, for distinguishing between resorption of  
CC non-mineralized and mineralized cartilage, and for measuring total  
CC cartilage resorption in vivo. The present sequence represents a linear  
CC synthetic peptide based on the carboxy-terminal telopeptide sequence of  
CC human collagen type II  
XX  
SQ Sequence 10 AA;  
Query Match 100.0%; Score 26; DB 4; Length 10;  
Best Local Similarity 100.0%; Pred. No. 43;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 EKAGG 5  
Db 3 EKAGG 7  
Search completed: September 18, 2004, 04:25:35  
Job time : 28.7647 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 18, 2004, 04:20:54 ; Search time 8.23529 Seconds  
(without alignments)  
31.344 Million cell updates/sec

Title: US-10-615-959-44  
Perfect score: 26  
Sequence: 1 EKAGG 5

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA.\*  
1: /cgn2\_6/ptodata/2/iaa/5A COMB.pap.\*  
2: /cgn2\_6/ptodata/2/iaa/5B COMB.pap.\*  
3: /cgn2\_6/ptodata/2/iaa/6A COMB.pap.\*  
4: /cgn2\_6/ptodata/2/iaa/6B COMB.pap.\*  
5: /cgn2\_6/ptodata/2/iaa/PCTUS COMB.pap.\*  
6: /cgn2\_6/ptodata/2/iaa/backfiles1.pap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	26	100.0	5	4	US-10-009-999A-44
2	26	100.0	6	4	US-10-009-999A-39
3	26	100.0	7	3	US-08-963-825-15
4	26	100.0	7	3	US-08-817-397B-16
5	26	100.0	7	4	US-09-500-811-15
6	26	100.0	7	4	US-09-570-573-15
7	26	100.0	7	4	US-09-385-740B-21
8	26	100.0	7	4	US-09-548-608-15
9	26	100.0	7	4	US-09-714-146-16
10	26	100.0	7	4	US-10-073-679-21
11	26	100.0	7	4	US-10-009-999A-21
12	26	100.0	8	4	US-09-385-740B-22
13	26	100.0	8	4	US-09-385-740B-24
14	26	100.0	8	4	US-10-073-679-22
15	26	100.0	8	4	US-10-073-679-24
16	26	100.0	8	4	US-10-009-999A-22
17	26	100.0	8	4	US-10-009-999A-24
18	26	100.0	8	4	US-10-009-999A-40
19	26	100.0	9	4	US-09-385-740B-23
20	26	100.0	9	4	US-09-385-740B-25
21	26	100.0	9	4	US-10-073-679-23
22	26	100.0	9	4	US-10-073-679-25
23	26	100.0	9	4	US-10-009-999A-23
24	26	100.0	9	4	US-10-009-999A-25
25	26	100.0	9	4	US-10-009-999A-41
26	26	100.0	10	4	US-09-385-740B-26
27	26	100.0	10	4	US-10-073-679-26

28	26	100.0	10	4	US-10-009-999A-26	Sequence 26, Appl
29	26	100.0	11	4	US-09-385-740B-31	Sequence 31, Appl
30	26	100.0	11	4	US-10-073-679-31	Sequence 31, Appl
31	26	100.0	11	4	US-10-009-999A-31	Sequence 31, Appl
32	26	100.0	11	4	US-10-009-999A-42	Sequence 42, Appl
33	26	100.0	12	4	US-09-385-740B-30	Sequence 30, Appl
34	26	100.0	12	4	US-10-073-679-30	Sequence 30, Appl
35	26	100.0	12	4	US-10-009-999A-30	Sequence 30, Appl
36	26	100.0	12	4	US-10-009-999A-43	Sequence 43, Appl
37	26	100.0	13	4	US-09-385-740B-27	Sequence 27, Appl
38	26	100.0	13	4	US-10-073-679-27	Sequence 27, Appl
39	26	100.0	13	4	US-10-009-999A-27	Sequence 27, Appl
40	26	100.0	15	4	US-08-278-774-20	Sequence 20, Appl
41	26	100.0	16	4	US-09-385-740B-28	Sequence 28, Appl
42	26	100.0	16	4	US-10-073-679-28	Sequence 28, Appl
43	26	100.0	16	4	US-10-009-999A-28	Sequence 28, Appl
44	26	100.0	17	4	US-09-385-740B-29	Sequence 29, Appl
45	26	100.0	17	4	US-10-073-679-29	Sequence 29, Appl

## ALIGNMENTS

RESULT 1  
US-10-009-999A-44  
; Sequence 44, Application US/10009999A  
; Patent No. 6602980  
; GENERAL INFORMATION:  
; APPLICANT: Eyre, David R.  
; TITLE OF INVENTION: SYNTHETIC PEPTIDES FOR COLLAGEN RESORPTION ASSAYS  
; FILE REFERENCE: WPOS-1-18220  
; CURRENT APPLICATION NUMBER: US/10/009,999A  
; CURRENT FILING DATE: 2003-01-22  
; PRIOR APPLICATION NUMBER: PCT/US99/29357  
; PRIOR FILING DATE: 1999-12-10  
; PRIOR APPLICATION NUMBER: US 09/335,098  
; PRIOR FILING DATE: 1999-06-17  
; PRIOR APPLICATION NUMBER: US 60/141,574  
; PRIOR FILING DATE: 1999-06-29  
; PRIOR APPLICATION NUMBER: US 60/142,274  
; PRIOR FILING DATE: 1999-07-02  
; PRIOR APPLICATION NUMBER: US 60/142,675  
; PRIOR FILING DATE: 1999-07-07  
; PRIOR APPLICATION NUMBER: US 09/385,740  
; PRIOR FILING DATE: 1999-08-30  
; NUMBER OF SEQ ID NOS: 45  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 44  
; LENGTH: 5  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: MISC FEATURE  
; OTHER INFORMATION: synthetic peptide corresponding to carboxy-terminal  
; OTHER INFORMATION: telopeptide sequence of human type III collagen  
US-10-009-999A-44

Query Match 100.0%; Score 26; DB 4; Length 5;  
Best Local Similarity 100.0%; Pred. No. 3e+05;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 EKAGG 5  
Db 1 EKAGG 5

RESULT 2  
US-10-009-999A-39  
; Sequence 39, Application US/10009999A  
; Patent No. 6602980  
; GENERAL INFORMATION:  
; APPLICANT: Eyre, David R.  
; TITLE OF INVENTION: SYNTHETIC PEPTIDES FOR COLLAGEN RESORPTION ASSAYS

FILE REFERENCE: WROS-1-18220  
CURRENT APPLICATION NUMBER: US/10/009,999A  
PRIOR FILING DATE: 2003-01-22  
PRIOR APPLICATION NUMBER: PCT/US99/29357  
PRIOR FILING DATE: 1999-12-10  
PRIOR APPLICATION NUMBER: US 09/335,098  
PRIOR FILING DATE: 1999-06-17  
PRIOR APPLICATION NUMBER: US 60/141,574  
PRIOR FILING DATE: 1999-06-29  
PRIOR APPLICATION NUMBER: US 60/142,274  
PRIOR FILING DATE: 1999-07-02  
PRIOR APPLICATION NUMBER: US 60/142,675  
PRIOR FILING DATE: 1999-07-07  
PRIOR APPLICATION NUMBER: US 09/385,740  
PRIOR FILING DATE: 1999-08-30  
NUMBER OF SEQ ID NOS: 45  
SOFTWARE: PatentIn version 3.2  
SEQ ID NO 39  
LENGTH: 6  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: MISC FEATURE  
OTHER INFORMATION: synthetic peptide corresponding to carboxy-terminal  
OTHER INFORMATION: telopeptide sequence of human type III collagen  
US-10-009-999A-39

Query Match 100.0%; Score 26; DB 4; Length 6;  
Best Local Similarity 100.0%; Pred. No. 3e+05;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EXAGG 5  
Db 1 EXAGG 5

RESULT 3  
US-08-963-825-15  
Sequence 15, Application US/08963825  
Patent No. 6110689  
GENERAL INFORMATION:  
APPLICANT: Qvist, Per  
TITLE OF INVENTION: A Method for Assaying Collagen Fragments  
TITLE OF INVENTION: in Body Fluids, A Test Kit and Means for Carrying Out the  
TITLE OF INVENTION: Method and Use of the Method to Diagnose the Presence of  
TITLE OF INVENTION: Disorders Associated with the Metabolism of  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Darby & Darby PC  
STREET: 805 Third Avenue  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10022  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/963,825  
FILING DATE:  
CLASSIFICATION: 436  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/187,319  
FILING DATE: 21-JAN-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Gogoris, Adda C  
REGISTRATION NUMBER: 29,714  
REFERENCE/DOCKET NUMBER: 4305/08701  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-527-7700

Query Match 100.0%; Score 26; DB 4; Length 6;  
Best Local Similarity 100.0%; Pred. No. 3e+05;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EXAGG 5  
Db 1 EXAGG 5

RESULT 4  
US-08-817-397B-16  
Sequence 16, Application US/08817397B  
Patent No. 6210902  
GENERAL INFORMATION:  
APPLICANT: Qvist, Per  
APPLICANT: BONDE, Martin  
TITLE OF INVENTION: ESTIMATION OF THE FRAGMENTATION  
TITLE OF INVENTION: PATTERN OF COLLAGEN IN BODY FLUIDS AND THE DIAGNOSIS  
TITLE OF INVENTION: OF DISORDERS...  
NUMBER OF SEQUENCES: 20  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pennie & Edmonds LLP  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: NY  
COUNTRY: USA  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/817,397B  
FILING DATE: 11-JUN-1997  
CLASSIFICATION: 436  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/EP95/04055  
FILING DATE: 16-OCT-1995  
APPLICATION NUMBER: DK 1194/94  
FILING DATE: 17-OCT-1994  
APPLICATION NUMBER: GB 9506050  
FILING DATE: 24-MAR-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Lawrence, Iii, Stanton T  
REGISTRATION NUMBER: 25,736  
REFERENCE/DOCKET NUMBER: 8969-008-999  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-790-9090  
TELEFAX: 212-863-9741  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 16:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 7 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide

TELEFAX: 212-753-6237  
TELEX: 236687  
INFORMATION FOR SEQ ID NO: 15:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 7 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FRAGMENT TYPE: C-terminal  
ORIGINAL SOURCE:  
ORGANISM: Homo sapiens  
IMMEDIATE SOURCE:  
CLONE: COLLAGEN TYPE III - ALPHA 1 - C TERM  
US-08-963-825-15

Query Match 100.0%; Score 26; DB 3; Length 7;  
Best Local Similarity 100.0%; Pred. No. 3e+05;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EXAGG 5  
Db 1 EXAGG 5

RESULT 4  
US-08-817-397B-16  
Sequence 16, Application US/08817397B  
Patent No. 6210902  
GENERAL INFORMATION:  
APPLICANT: Qvist, Per  
APPLICANT: BONDE, Martin  
TITLE OF INVENTION: ESTIMATION OF THE FRAGMENTATION  
TITLE OF INVENTION: PATTERN OF COLLAGEN IN BODY FLUIDS AND THE DIAGNOSIS  
TITLE OF INVENTION: OF DISORDERS...  
NUMBER OF SEQUENCES: 20  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pennie & Edmonds LLP  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: NY  
COUNTRY: USA  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/817,397B  
FILING DATE: 11-JUN-1997  
CLASSIFICATION: 436  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/EP95/04055  
FILING DATE: 16-OCT-1995  
APPLICATION NUMBER: DK 1194/94  
FILING DATE: 17-OCT-1994  
APPLICATION NUMBER: GB 9506050  
FILING DATE: 24-MAR-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Lawrence, Iii, Stanton T  
REGISTRATION NUMBER: 25,736  
REFERENCE/DOCKET NUMBER: 8969-008-999  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-790-9090  
TELEFAX: 212-863-9741  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 16:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 7 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide



US-08-817-397B-16

Query Match 100.0%; Score 26; DB 3; Length 7;  
Best Local Similarity 100.0%; Pred. No. 3e+05;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 EKAGG 5  
|||||  
Db 1 EKAGG 5

RESULT 5

US-09-500-811-15  
; Sequence 15, Application US/09500811  
; Patent No. 632314  
; GENERAL INFORMATION:

; APPLICANT: Qvist, Per  
; APPLICANT: Bonde, Martin  
; TITLE OF INVENTION: A Method for Assaying Collagen Fragments  
; TITLE OF INVENTION: in Body Fluids, A Test Kit and Means for Carrying Out the  
; TITLE OF INVENTION: Method and Use of the Method to Diagnose the Presence of  
; TITLE OF INVENTION: Disorders Associated with the Metabolism of  
; NUMBER OF SEQUENCES: 21  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Darby & Darby PC  
; STREET: 805 Third Avenue  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10022

; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/500.811  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/187,319  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Gogoris, Adda C  
; REGISTRATION NUMBER: 29,714  
; REFERENCE/DOCKET NUMBER: 4305/08701  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212-527-7700  
; TELEFAX: 212-753-6237  
; TELEX: 236687

; INFORMATION FOR SEQ ID NO: 15:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 7 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; FRAGMENT TYPE: C-terminal  
; ORIGINAL SOURCE:  
; ORGANISM: Homo sapiens  
; IMMEDIATE SOURCE:  
; CLONE: COLLAGEN TYPE III - ALPHA 1 - C TERM

US-09-500-811-15  
; Sequence 15, Application US/09500811  
; Patent No. 632314  
; GENERAL INFORMATION:  
; APPLICANT: Qvist, Per  
; APPLICANT: Bonde, Martin  
; TITLE OF INVENTION: A Method for Assaying Collagen Fragments  
; TITLE OF INVENTION: in Body Fluids, A Test Kit and Means for Carrying Out the  
; TITLE OF INVENTION: Method and Use of the Method to Diagnose the Presence of  
; TITLE OF INVENTION: Disorders Associated with the Metabolism of  
; NUMBER OF SEQUENCES: 21  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Darby & Darby PC  
; STREET: 805 Third Avenue  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10022

Query Match 100.0%; Score 26; DB 4; Length 7;  
Best Local Similarity 100.0%; Pred. No. 3e+05;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 EKAGG 5  
|||||  
Db 1 EKAGG 5

RESULT 6

US-09-570-573-15

; Sequence 15, Application US/09570573  
; Patent No. 6342361  
; GENERAL INFORMATION:

; APPLICANT: Qvist, Per  
; APPLICANT: Bonde, Martin  
; TITLE OF INVENTION: A Method for Assaying Collagen Fragments  
; TITLE OF INVENTION: in Body Fluids, A Test Kit and Means for Carrying Out the  
; TITLE OF INVENTION: Method and Use of the Method to Diagnose the Presence of  
; TITLE OF INVENTION: Disorders Associated with the Metabolism of  
; NUMBER OF SEQUENCES: 21  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Darby & Darby PC  
; STREET: 805 Third Avenue  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10022

; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/570,573  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/187,319  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Gogoris, Adda C  
; REGISTRATION NUMBER: 29,714  
; REFERENCE/DOCKET NUMBER: 4305/08701  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212-527-7700  
; TELEFAX: 212-753-6237  
; TELEX: 236687

; INFORMATION FOR SEQ ID NO: 15:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 7 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; FRAGMENT TYPE: C-terminal  
; ORIGINAL SOURCE:  
; ORGANISM: Homo sapiens  
; IMMEDIATE SOURCE:  
; CLONE: COLLAGEN TYPE III - ALPHA 1 - C TERM

US-09-570-573-15

Query Match 100.0%; Score 26; DB 4; Length 7;  
Best Local Similarity 100.0%; Pred. No. 3e+05;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 EKAGG 5  
|||||  
Db 1 EKAGG 5

RESULT 7

US-09-385-740B-21  
; Sequence 21, Application US/09385740B  
; Patent No. 6348320  
; GENERAL INFORMATION:

; APPLICANT: Eyre, David  
; TITLE OF INVENTION: CARTILAGE RESORPTION ASSAYS  
; FILE REFERENCE: WROS-1-14269  
; CURRENT APPLICATION NUMBER: US/09/385,740B  
; CURRENT FILING DATE: 1999-08-30  
; PRIOR APPLICATION NUMBER: US 60/142,274  
; PRIOR FILING DATE: 1999-07-02  
; PRIOR APPLICATION NUMBER: US 60/141,574

```

; PRIOR FILING DATE: 1999-06-29
; PRIOR APPLICATION NUMBER: US 09/335,098
; PRIOR FILING DATE: 1999-06-17
; PRIOR APPLICATION NUMBER: US 60/089,823
; PRIOR FILING DATE: 1998-06-19
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 21
; LENGTH: 7
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: PEPTIDE
; LOCATION: (1)..(7)
; OTHER INFORMATION: syn- corr. to N-terminal telopeptide seq. of hu. type II collagen
US-09-385-740B-21

Query Match          100.0%; Score 26; DB 4; Length 7;
Best Local Similarity 100.0%; Pred. No. 3e+05; 0; Gaps 0;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 EKAGG 5
Db      2 EKAGG 6

RESULT 8
US-09-548-608-15
; Sequence 15, Application US/09548608
; Patent No. 6355442
; GENERAL INFORMATION:
; APPLICANT: Qvist, Per
; APPLICANT: Bonde, Martin
; TITLE OF INVENTION: A Method for Assaying Collagen Fragments
; TITLE OF INVENTION: In Body Fluids, A Test Kit and Means for Carrying Out the
; TITLE OF INVENTION: Method and Use of the Method to Diagnose the Presence of
; TITLE OF INVENTION: Disorders Associated with the Metabolism of
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Darby & Darby PC
; STREET: 805 Third Avenue
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/548,608
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/187,319
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Gogorie, Adda C
; REGISTRATION NUMBER: 29,714
; REFERENCE/DOCKET NUMBER: 4305/08701
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-527-7700
; TELEFAX: 212-753-6237
; TELEX: 236687
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FRAGMENT TYPE: C-terminal
; ORIGINAL SOURCE:

```

```

; ORGANISM: Homo sapiens
; IMMEDIATE SOURCE:
; CLONE: COLLAGEN TYPE III - ALPHA 1 - C TERM
US-09-548-608-15

Query Match          100.0%; Score 26; DB 4; Length 7;
Best Local Similarity 100.0%; Pred. No. 3e+05; 0; Indels 0; Gaps 0;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 EKAGG 5
Db      1 EKAGG 5

RESULT 9
US-09-714-146-16
; Sequence 16, Application US/09714146
; Patent No. 6372442
; GENERAL INFORMATION:
; APPLICANT: Per Ovist
; APPLICANT: Martin Bonde
; TITLE OF INVENTION: A Method of Characterizing the Degradation of Type II Collagen
; FILE REFERENCE: 8969-020-999
; CURRENT APPLICATION NUMBER: US/09/714,146
; CURRENT FILING DATE: 2001-11-17
; PRIOR APPLICATION NUMBER: 08/817,397
; PRIOR FILING DATE: 1997-06-11
; PRIOR APPLICATION NUMBER: PCT/EP95/04055
; PRIOR FILING DATE: 1995-10-16
; PRIOR APPLICATION NUMBER: 9506050
; PRIOR FILING DATE: 1995-03-24
; PRIOR APPLICATION NUMBER: 1194/94
; PRIOR FILING DATE: 1994-10-17
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16
; LENGTH: 7
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-714-146-16

Query Match          100.0%; Score 26; DB 4; Length 7;
Best Local Similarity 100.0%; Pred. No. 3e+05; 0; Indels 0; Gaps 0;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 EKAGG 5
Db      1 EKAGG 5

RESULT 10
US-10-073-679-21
; Sequence 21, Application US/10073679
; Patent No. 6566492
; GENERAL INFORMATION:
; APPLICANT: Eyre, David
; TITLE OF INVENTION: CARTILAGE RESORPTION ASSAYS
; FILE REFERENCE: WROS-1-14269
; CURRENT APPLICATION NUMBER: US/10/073,679
; CURRENT FILING DATE: 2002-02-11
; PRIOR APPLICATION NUMBER: 09/385,740
; PRIOR FILING DATE: 1999-08-30
; PRIOR APPLICATION NUMBER: US 60/142,274
; PRIOR FILING DATE: 1999-07-02
; PRIOR APPLICATION NUMBER: US 60/141,574
; PRIOR FILING DATE: 1999-06-29
; PRIOR APPLICATION NUMBER: US 09/335,098
; PRIOR FILING DATE: 1999-06-17
; PRIOR APPLICATION NUMBER: US 60/089,823
; PRIOR FILING DATE: 1998-06-19
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 21

```

; LENGTH: 7  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: PEPTIDE  
; LOCATION: (1)..(7)  
; OTHER INFORMATION: syn- corr. to N-terminal telopeptide seq. of hu. type II collagen  
US-10-073-679-21

Query Match 100.0%; Score 26; DB 4; Length 7;  
Best Local Similarity 100.0%; Pred. No. 3e+05;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 EKAGG 5  
Db 2 EKAGG 6

## RESULT 11

US-10-009-999A-21  
; Sequence 21, Application US/10009999A  
; Patent No. 6602980  
; GENERAL INFORMATION:  
; APPLICANT: Eyre, David R.  
; TITLE OF INVENTION: SYNTHETIC PEPTIDES FOR COLLAGEN RESORPTION ASSAYS  
; FILE REFERENCE: WROS-1-18220  
; CURRENT APPLICATION NUMBER: US/10/009,999A  
; CURRENT FILING DATE: 2003-01-22  
; PRIOR APPLICATION NUMBER: PCT/US99/29357  
; PRIOR FILING DATE: 1999-12-10  
; PRIOR APPLICATION NUMBER: US 09/335,098  
; PRIOR FILING DATE: 1999-06-17  
; PRIOR APPLICATION NUMBER: US 60/141,574  
; PRIOR FILING DATE: 1999-06-29  
; PRIOR APPLICATION NUMBER: US 60/142,274  
; PRIOR FILING DATE: 1999-07-02  
; PRIOR APPLICATION NUMBER: US 60/142,675  
; PRIOR FILING DATE: 1999-07-07  
; PRIOR APPLICATION NUMBER: US 09/385,740  
; PRIOR FILING DATE: 1999-08-30  
; NUMBER OF SEQ ID NOS: 45  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 21  
; LENGTH: 7  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; NAME/KEY: MISC FEATURE  
; OTHER INFORMATION: synthetic peptide corresponding to amino-terminal  
; OTHER INFORMATION: telopeptide sequence of human type II collagen  
US-10-009-999A-21

Query Match 100.0%; Score 26; DB 4; Length 7;  
Best Local Similarity 100.0%; Pred. No. 3e+05;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 EKAGG 5  
Db 2 EKAGG 6

## RESULT 12

US-09-385-740B-22  
; Sequence 22, Application US/09385740B  
; Patent No. 6348320  
; GENERAL INFORMATION:  
; APPLICANT: Eyre, David  
; TITLE OF INVENTION: CARTILAGE RESORPTION ASSAYS  
; FILE REFERENCE: WROS-1-14269  
; CURRENT APPLICATION NUMBER: US/09/385,740B  
; CURRENT FILING DATE: 1999-08-30  
; PRIOR APPLICATION NUMBER: US 60/142,274  
; PRIOR FILING DATE: 1999-07-02

; PRIOR APPLICATION NUMBER: US 60/141,574  
; PRIOR FILING DATE: 1999-06-29  
; PRIOR APPLICATION NUMBER: US 09/335,098  
; PRIOR FILING DATE: 1999-06-17  
; PRIOR APPLICATION NUMBER: US 60/089,823  
; PRIOR FILING DATE: 1998-06-19  
; NUMBER OF SEQ ID NOS: 31  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 22  
; LENGTH: 8  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: PEPTIDE  
; LOCATION: (1)..(8)  
; OTHER INFORMATION: syn- corr. to N-terminal telopeptide seq. of hu. type II collagen  
US-09-385-740B-22

Query Match 100.0%; Score 26; DB 4; Length 8;  
Best Local Similarity 100.0%; Pred. No. 3e+05;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 EKAGG 5  
Db 2 EKAGG 6

## RESULT 13

US-09-385-740B-24  
; Sequence 24, Application US/09385740B  
; Patent No. 6348320  
; GENERAL INFORMATION:  
; APPLICANT: Eyre, David  
; TITLE OF INVENTION: CARTILAGE RESORPTION ASSAYS  
; FILE REFERENCE: WROS-1-14269  
; CURRENT APPLICATION NUMBER: US/09/385,740B  
; CURRENT FILING DATE: 1999-08-30  
; PRIOR APPLICATION NUMBER: US 60/142,274  
; PRIOR FILING DATE: 1999-07-02  
; PRIOR APPLICATION NUMBER: US 60/141,574  
; PRIOR FILING DATE: 1999-06-29  
; PRIOR APPLICATION NUMBER: US 09/335,098  
; PRIOR FILING DATE: 1999-06-17  
; PRIOR APPLICATION NUMBER: US 60/089,823  
; PRIOR FILING DATE: 1998-06-19  
; NUMBER OF SEQ ID NOS: 31  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 24  
; LENGTH: 8  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: PEPTIDE  
; LOCATION: (1)..(8)  
; OTHER INFORMATION: syn- corr. to N-terminal telopeptide seq. of hu. type II collagen  
US-09-385-740B-24

Query Match 100.0%; Score 26; DB 4; Length 8;  
Best Local Similarity 100.0%; Pred. No. 3e+05;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 EKAGG 5  
Db 3 EKAGG 7

## RESULT 14

US-10-073-679-22  
; Sequence 22, Application US/10073679  
; Patent No. 6566492  
; GENERAL INFORMATION:  
; APPLICANT: Eyre, David  
; TITLE OF INVENTION: CARTILAGE RESORPTION ASSAYS

FILE REFERENCE: WROS-1-14269  
 CURRENT APPLICATION NUMBER: US/10/073,679  
 CURRENT FILING DATE: 2002-02-11  
 PRIOR APPLICATION NUMBER: 09/385,740  
 PRIOR FILING DATE: 1999-08-30  
 PRIOR APPLICATION NUMBER: US 60/142,274  
 PRIOR FILING DATE: 1999-07-02  
 PRIOR APPLICATION NUMBER: US 60/141,574  
 PRIOR FILING DATE: 1999-06-29  
 PRIOR APPLICATION NUMBER: US 09/335,098  
 PRIOR FILING DATE: 1999-06-17  
 PRIOR APPLICATION NUMBER: US 60/089,823  
 PRIOR FILING DATE: 1998-06-19  
 NUMBER OF SEQ ID NOS: 31  
 SOFTWARE: PatentIn version 3.0  
 SEQ ID NO 22  
 LENGTH: 8  
 TYPE: PRT  
 ORGANISM: Homo sapiens  
 FEATURE:  
 NAME/KEY: PEPTIDE  
 LOCATION: (1)..(8)  
 OTHER INFORMATION: syn- corr. to N-terminal telopeptide seq. of hu. type II collagen

Search completed: September 18, 2004, 04:32:23  
 Job time : 8.23529 secs

US-10-073-679-22

Query Match 100.0%; Score 26; DB 4; Length 8;  
 Best Local Similarity 100.0%; Pred. No. 3e+05; Indels 0; Gaps 0;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EKAGG 5  
 Db 2 EKAGG 6

RESULT 15

US-10-073-679-24  
 Sequence 24, Application US/10073679  
 Patent No. 6566492  
 GENERAL INFORMATION:  
 APPLICANT: Eyre, David  
 TITLE OF INVENTION: CARTILAGE RESORPTION ASSAYS  
 FILE REFERENCE: WROS-1-14269  
 CURRENT APPLICATION NUMBER: US/10/073,679  
 CURRENT FILING DATE: 2002-02-11  
 PRIOR APPLICATION NUMBER: 09/385,740  
 PRIOR FILING DATE: 1999-08-30  
 PRIOR APPLICATION NUMBER: US 60/142,274  
 PRIOR FILING DATE: 1999-07-02  
 PRIOR APPLICATION NUMBER: US 60/141,574  
 PRIOR FILING DATE: 1999-06-29  
 PRIOR APPLICATION NUMBER: US 09/335,098  
 PRIOR FILING DATE: 1999-06-17  
 PRIOR APPLICATION NUMBER: US 60/089,823  
 PRIOR FILING DATE: 1998-06-19  
 NUMBER OF SEQ ID NOS: 31  
 SOFTWARE: PatentIn version 3.0  
 SEQ ID NO 24  
 LENGTH: 8  
 TYPE: PRT  
 ORGANISM: Homo sapiens  
 FEATURE:  
 NAME/KEY: PEPTIDE  
 LOCATION: (1)..(8)  
 OTHER INFORMATION: syn- corr. to N-terminal telopeptide seq. of hu. type II collagen

Query Match 100.0%; Score 26; DB 4; Length 8;  
 Best Local Similarity 100.0%; Pred. No. 3e+05; Indels 0; Gaps 0;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EKAGG 5  
 Db 3 EKAGG 7

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 18, 2004, 04:29:56 ; Search time 23.7255 Seconds  
(without alignments)  
67.677 Million cell updates/sec

Title: US-10-615-959-44  
Perfect score: 26  
Sequence: 1 BKAGG 5

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1342398 seqs, 321133274 residues

Total number of hits satisfying chosen parameters: 1342398

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database : Published Applications AA:  
1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*  
2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*  
8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/2/pubpaa/US09A\_PUBCOMB.pep.\*  
10: /cgn2\_6/ptodata/2/pubpaa/US09B\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/2/pubpaa/US09C\_PUBCOMB.pep.\*  
12: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep.\*  
13: /cgn2\_6/ptodata/2/pubpaa/US10A\_PUBCOMB.pep.\*  
14: /cgn2\_6/ptodata/2/pubpaa/US10B\_PUBCOMB.pep.\*  
15: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pep.\*  
16: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep.\*  
17: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*  
18: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	26	100.0	5	12	US-10-615-959-44
2	26	100.0	6	12	US-10-615-959-44
3	26	100.0	7	12	US-10-615-959-39
4	26	100.0	7	13	US-10-615-959-21
5	26	100.0	7	13	US-10-073-679-21
6	26	100.0	7	14	US-10-058-124-15
7	26	100.0	7	14	US-10-408-178-21
8	26	100.0	8	12	US-10-615-959-22
9	26	100.0	8	12	US-10-615-959-24
10	26	100.0	8	13	US-10-615-959-40
11	26	100.0	8	13	US-10-073-679-22
12	26	100.0	8	13	US-10-073-679-24
13	26	100.0	8	14	US-10-408-178-22
14	26	100.0	9	12	US-10-615-959-23
15	26	100.0	9	12	US-10-615-959-25

16	26	100.0	9	12	US-10-615-959-41
17	26	100.0	9	13	US-10-073-679-23
18	26	100.0	9	13	US-10-073-679-25
19	26	100.0	9	14	US-10-408-178-23
20	26	100.0	9	14	US-10-408-178-25
21	26	100.0	10	9	US-09-848-967-22
22	26	100.0	10	12	US-10-615-959-26
23	26	100.0	10	13	US-10-073-679-26
24	26	100.0	10	14	US-10-408-178-26
25	26	100.0	11	12	US-10-615-959-31
26	26	100.0	11	12	US-10-615-959-42
27	26	100.0	11	13	US-10-073-679-31
28	26	100.0	11	14	US-10-408-178-31
29	26	100.0	12	12	US-10-615-959-30
30	26	100.0	12	12	US-10-615-959-43
31	26	100.0	12	13	US-10-073-679-30
32	26	100.0	12	14	US-10-408-178-30
33	26	100.0	12	14	US-10-615-959-27
34	26	100.0	13	12	US-10-073-679-27
35	26	100.0	13	13	US-10-408-178-27
36	26	100.0	14	14	US-10-145-012-4
37	26	100.0	16	9	US-09-815-837-120
38	26	100.0	16	12	US-10-615-959-28
39	26	100.0	16	13	US-10-073-679-28
40	26	100.0	16	14	US-10-408-178-28
41	26	100.0	17	9	US-09-815-837-123
42	26	100.0	17	12	US-10-615-959-29
43	26	100.0	17	13	US-10-073-679-29
44	26	100.0	17	14	US-10-408-178-29
45	26	100.0	19	15	US-10-366-125-4

## ALIGNMENTS

RESULT 1  
US-10-615-959-44  
; Sequence 44, Application US/10615959  
; Publication No. US20040048321A1  
; GENERAL INFORMATION:  
; APPLICANT: Eyre, David R.  
; TITLE OF INVENTION: SYNTHETIC PEPTIDES FOR COLLAGEN RESORPTION ASSAYS  
; FILE REFERENCE: WROS-1-18220  
; CURRENT APPLICATION NUMBER: US/10/615,959  
; CURRENT FILING DATE: 2003-07-08  
; PRIOR APPLICATION NUMBER: US/10/009,999A  
; PRIOR FILING DATE: 2003-01-22  
; PRIOR APPLICATION NUMBER: PCT/US99/29357  
; PRIOR FILING DATE: 1999-12-10  
; PRIOR APPLICATION NUMBER: US 09/335,098  
; PRIOR FILING DATE: 1999-06-17  
; PRIOR APPLICATION NUMBER: US 60/141,574  
; PRIOR FILING DATE: 1999-06-29  
; PRIOR APPLICATION NUMBER: US 60/142,274  
; PRIOR FILING DATE: 1999-07-02  
; PRIOR APPLICATION NUMBER: US 60/142,675  
; PRIOR FILING DATE: 1999-07-07  
; PRIOR APPLICATION NUMBER: US 09/385,740  
; NUMBER OF SEQ ID NOS: 45  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 44  
; LENGTH: 5  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: MISC FEATURE  
; OTHER INFORMATION: synthetic peptide corresponding to carboxy-terminal  
; OTHER INFORMATION: telopeptide sequence of human type III collagen  
US-10-615-959-44

Query Match 100.0%; Score 26; DB 12; Length 5;  
Best Local Similarity 100.0%; Pred. No. 1.2e+06;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 EKAGG 5  
Db 1 EKAGG 5

RESULT 2  
US-10-615-959-39  
; Sequence 39, Application US/10615959  
; Publication No. US20040048321A1  
; GENERAL INFORMATION:  
; APPLICANT: Eyre, David R.  
; TITLE OF INVENTION: SYNTHETIC PEPTIDES FOR COLLAGEN RESORPTION ASSAYS  
; FILE REFERENCE: WROS-1-18220  
; CURRENT APPLICATION NUMBER: US/10/615,959  
; CURRENT FILING DATE: 2003-07-08  
; PRIOR APPLICATION NUMBER: US/10/009,999A  
; PRIOR FILING DATE: 2003-01-22  
; PRIOR APPLICATION NUMBER: PCT/US99/29357  
; PRIOR FILING DATE: 1999-12-10  
; PRIOR APPLICATION NUMBER: US 09/335,098  
; PRIOR FILING DATE: 1999-06-17  
; PRIOR APPLICATION NUMBER: US 60/141,574  
; PRIOR FILING DATE: 1999-06-29  
; PRIOR APPLICATION NUMBER: US 60/142,274  
; PRIOR FILING DATE: 1999-07-02  
; PRIOR APPLICATION NUMBER: US 60/142,675  
; PRIOR FILING DATE: 1999-07-07  
; PRIOR APPLICATION NUMBER: US 09/385,740  
; PRIOR FILING DATE: 1999-08-30  
; NUMBER OF SEQ ID NOS: 45  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 39  
; LENGTH: 6  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: MISC FEATURE  
; OTHER INFORMATION: synthetic peptide corresponding to carboxy-terminal  
; OTHER INFORMATION: telopeptide sequence of human type III collagen

US-10-615-959-39

Query Match 100.0%; Score 26; DB 12; Length 6;  
Best Local Similarity 100.0%; Pred. No. 1.2e+06;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 EKAGG 5  
Db 1 EKAGG 5

RESULT 3  
US-10-615-959-21  
; Sequence 21, Application US/10615959  
; Publication No. US20040048321A1  
; GENERAL INFORMATION:  
; APPLICANT: Eyre, David R.  
; TITLE OF INVENTION: SYNTHETIC PEPTIDES FOR COLLAGEN RESORPTION ASSAYS  
; FILE REFERENCE: WROS-1-18220  
; CURRENT APPLICATION NUMBER: US/10/615,959  
; CURRENT FILING DATE: 2003-07-08  
; PRIOR APPLICATION NUMBER: US/10/009,999A  
; PRIOR FILING DATE: 2003-01-22  
; PRIOR APPLICATION NUMBER: PCT/US99/29357  
; PRIOR FILING DATE: 1999-12-10  
; PRIOR APPLICATION NUMBER: US 09/335,098  
; PRIOR FILING DATE: 1999-06-17  
; PRIOR APPLICATION NUMBER: US 60/141,574  
; PRIOR FILING DATE: 1999-06-29  
; PRIOR APPLICATION NUMBER: US 60/142,274  
; PRIOR FILING DATE: 1999-07-02  
; PRIOR APPLICATION NUMBER: US 60/142,675  
; PRIOR FILING DATE: 1999-07-07  
; PRIOR APPLICATION NUMBER: US 09/385,740  
; PRIOR FILING DATE: 1999-08-30  
; NUMBER OF SEQ ID NOS: 45  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 39  
; LENGTH: 6  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: MISC FEATURE  
; OTHER INFORMATION: synthetic peptide corresponding to carboxy-terminal  
; OTHER INFORMATION: telopeptide sequence of human type III collagen

US-10-615-959-39

Query Match 100.0%; Score 26; DB 12; Length 6;  
Best Local Similarity 100.0%; Pred. No. 1.2e+06;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 EKAGG 5  
Db 1 EKAGG 5

RESULT 4  
US-10-073-679-21  
; Sequence 21, Application US/10073679  
; Publication No. US20020182639A1  
; GENERAL INFORMATION:  
; APPLICANT: Eyre, David  
; TITLE OF INVENTION: CARTILAGE RESORPTION ASSAYS  
; FILE REFERENCE: WROS-1-14269  
; CURRENT APPLICATION NUMBER: US/10/073,679  
; CURRENT FILING DATE: 2002-02-11  
; PRIOR APPLICATION NUMBER: US 09/385,740  
; PRIOR FILING DATE: 1999-08-30  
; PRIOR APPLICATION NUMBER: US 60/142,274  
; PRIOR FILING DATE: 1999-07-02  
; PRIOR APPLICATION NUMBER: US 60/141,574  
; PRIOR FILING DATE: 1999-06-29  
; PRIOR APPLICATION NUMBER: US 09/335,098  
; PRIOR FILING DATE: 1999-06-17  
; PRIOR APPLICATION NUMBER: US 60/089,823  
; PRIOR FILING DATE: 1998-06-19  
; NUMBER OF SEQ ID NOS: 31  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 21  
; LENGTH: 7  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: PEPTIDE  
; LOCATION: (1)...(7)  
; OTHER INFORMATION: syn- corr. to N-terminal telopeptide seq. of hu. type II collagen

US-10-073-679-21

Query Match 100.0%; Score 26; DB 13; Length 7;  
Best Local Similarity 100.0%; Pred. No. 1.2e+06;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 EKAGG 5  
Db 2 EKAGG 6

PRIOR FILING DATE: 1999-07-07  
; PRIOR APPLICATION NUMBER: US 09/385,740  
; PRIOR FILING DATE: 1999-08-30  
; NUMBER OF SEQ ID NOS: 45  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 21  
; LENGTH: 7  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: MISC FEATURE  
; OTHER INFORMATION: synthetic peptide corresponding to amino-terminal  
; OTHER INFORMATION: telopeptide sequence of human type II collagen

US-10-615-959-21

Query Match 100.0%; Score 26; DB 12; Length 7;  
Best Local Similarity 100.0%; Pred. No. 1.2e+06;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 EKAGG 5  
Db 2 EKAGG 6

RESULT 5  
US-10-058-124-15  
; Sequence 15, Application US/10058124  
; Publication No. US20030119058A1  
; GENERAL INFORMATION:  
; APPLICANT: Qvist, Per  
; TITLE OF INVENTION: A Method for Assaying Collagen Fragments

US-10-058-124-15

Query Match 100.0%; Score 26; DB 13; Length 7;  
Best Local Similarity 100.0%; Pred. No. 1.2e+06;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 EKAGG 5  
Db 2 EKAGG 6



```
; Sequence 24, Application US/10615959
; Publication No. US20040048321A1
; GENERAL INFORMATION:
; APPLICANT: Eyre, David R.
; TITLE OF INVENTION: SYNTHETIC PEPTIDES FOR COLLAGEN RESORPTION ASSAYS
; FILE REFERENCE: WROS-1-18220
; CURRENT APPLICATION NUMBER: US/10/615,959
; CURRENT FILING DATE: 2003-07-08
; PRIOR APPLICATION NUMBER: US/10/009,999A
; PRIOR FILING DATE: 2003-01-22
; PRIOR APPLICATION NUMBER: PCT/US99/29357
; PRIOR FILING DATE: 1999-12-10
; PRIOR APPLICATION NUMBER: US 09/335,098
; PRIOR FILING DATE: 1999-06-17
; PRIOR APPLICATION NUMBER: US 60/141,574
; PRIOR FILING DATE: 1999-06-29
; PRIOR APPLICATION NUMBER: US 60/142,274
; PRIOR FILING DATE: 1999-07-02
; PRIOR APPLICATION NUMBER: US 60/142,675
; PRIOR FILING DATE: 1999-07-07
; PRIOR APPLICATION NUMBER: US 09/385,740
; PRIOR FILING DATE: 1999-08-30
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 24
; LENGTH: 8
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: MISC FEATURE
; OTHER INFORMATION: synthetic peptide corresponding to amino-terminal
; OTHER INFORMATION: telopeptide sequence of human type II collagen
US-10-615-959-24

Query Match      100.0%; Score 26; DB 12; Length 8;
Best Local Similarity 100.0%; Pred. No. 1.2e+06;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 EKAGG 5
Db      |||||
        3 EKAGG 7

RESULT 9
US-10-615-959-40
; Sequence 40, Application US/10615959
; Publication No. US20040048321A1
; GENERAL INFORMATION:
; APPLICANT: Eyre, David R.
; TITLE OF INVENTION: SYNTHETIC PEPTIDES FOR COLLAGEN RESORPTION ASSAYS
; FILE REFERENCE: WROS-1-18220
; CURRENT APPLICATION NUMBER: US/10/615,959
; CURRENT FILING DATE: 2003-07-08
; PRIOR APPLICATION NUMBER: US/10/009,999A
; PRIOR FILING DATE: 2003-01-22
; PRIOR APPLICATION NUMBER: PCT/US99/29357
; PRIOR FILING DATE: 1999-12-10
; PRIOR APPLICATION NUMBER: US 09/335,098
; PRIOR FILING DATE: 1999-06-17
; PRIOR APPLICATION NUMBER: US 60/141,574
; PRIOR FILING DATE: 1999-06-29
; PRIOR APPLICATION NUMBER: US 60/142,274
; PRIOR FILING DATE: 1999-07-02
; PRIOR APPLICATION NUMBER: US 60/142,675
; PRIOR FILING DATE: 1999-07-07
; PRIOR APPLICATION NUMBER: US 09/385,740
; PRIOR FILING DATE: 1999-08-30
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 40
; LENGTH: 8
; TYPE: PRT
; ORGANISM: Homo sapiens
```

```
; FEATURE:
; NAME/KEY: MISC FEATURE
; OTHER INFORMATION: synthetic peptide corresponding to carboxy-terminal
; OTHER INFORMATION: telopeptide sequence of human type III collagen
US-10-615-959-40

Query Match      100.0%; Score 26; DB 12; Length 8;
Best Local Similarity 100.0%; Pred. No. 1.2e+06;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 EKAGG 5
Db      |||||
        4 EKAGG 8

RESULT 10
US-10-073-679-22
; Sequence 22, Application US/10073679
; Publication No. US20020182639A1
; GENERAL INFORMATION:
; APPLICANT: Eyre, David
; TITLE OF INVENTION: CARTILAGE RESORPTION ASSAYS
; FILE REFERENCE: WROS-1-14269
; CURRENT APPLICATION NUMBER: US/10/073,679
; CURRENT FILING DATE: 2002-02-11
; PRIOR APPLICATION NUMBER: 09/385,740
; PRIOR FILING DATE: 1999-08-30
; PRIOR APPLICATION NUMBER: US 60/142,274
; PRIOR FILING DATE: 1999-07-02
; PRIOR APPLICATION NUMBER: US 60/141,574
; PRIOR FILING DATE: 1999-06-29
; PRIOR APPLICATION NUMBER: US 09/335,098
; PRIOR FILING DATE: 1999-06-17
; PRIOR APPLICATION NUMBER: US 60/089,823
; PRIOR FILING DATE: 1998-06-19
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 22
; LENGTH: 8
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: PEPTIDE
; LOCATION: (1)..(8)
; OTHER INFORMATION: syn- corr. to N-terminal telopeptide seq. of hu. type II collagen
US-10-073-679-22

Query Match      100.0%; Score 26; DB 13; Length 8;
Best Local Similarity 100.0%; Pred. No. 1.2e+06;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 EKAGG 5
Db      |||||
        2 EKAGG 6

RESULT 11
US-10-073-679-24
; Sequence 24, Application US/10073679
; Publication No. US20020182639A1
; GENERAL INFORMATION:
; APPLICANT: Eyre, David
; TITLE OF INVENTION: CARTILAGE RESORPTION ASSAYS
; FILE REFERENCE: WROS-1-14269
; CURRENT APPLICATION NUMBER: US/10/073,679
; CURRENT FILING DATE: 2002-02-11
; PRIOR APPLICATION NUMBER: 09/385,740
; PRIOR FILING DATE: 1999-08-30
; PRIOR APPLICATION NUMBER: US 60/142,274
; PRIOR FILING DATE: 1999-07-02
; PRIOR APPLICATION NUMBER: US 60/141,574
; PRIOR FILING DATE: 1999-06-29
; PRIOR APPLICATION NUMBER: US 09/335,098
```



```
; PRIOR FILING DATE: 1999-06-17
; PRIOR APPLICATION NUMBER: US 60/089,823
; PRIOR FILING DATE: 1998-06-19
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 24
; LENGTH: 8
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: PEPTIDE
; LOCATION: (1)..(8)
; OTHER INFORMATION: syn- corr. to N-terminal telopeptide seq. of hu. type II collagen
US-10-073-679-24
```

```
Query Match          100.0%; Score 26; DB 13; Length 8;
Best Local Similarity 100.0%; Pred. No. 1.2e+06;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 EKAGG 5
    |||||
Db 3 EKAGG 7
```

```
RESULT 12
; Sequence 22, Application US/10408178
; Publication No. US20030166032A1
; GENERAL INFORMATION:
; APPLICANT: Eyre, David
; TITLE OF INVENTION: CARTILAGE RESORPTION ASSAYS
; FILE REFERENCE: WROS-1-14269
; CURRENT APPLICATION NUMBER: US/10/408,178
; CURRENT FILING DATE: 2003-04-03
; PRIOR APPLICATION NUMBER: US/10/073,679
; PRIOR FILING DATE: 2002-02-11
; PRIOR APPLICATION NUMBER: 09/385,740
; PRIOR FILING DATE: 1999-08-30
; PRIOR APPLICATION NUMBER: US 60/142,274
; PRIOR FILING DATE: 1999-07-02
; PRIOR APPLICATION NUMBER: US 60/141,574
; PRIOR FILING DATE: 1999-06-29
; PRIOR APPLICATION NUMBER: US 09/335,098
; PRIOR FILING DATE: 1999-06-17
; PRIOR APPLICATION NUMBER: US 60/089,823
; PRIOR FILING DATE: 1998-06-19
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 22
; LENGTH: 8
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: PEPTIDE
; LOCATION: (1)..(8)
; OTHER INFORMATION: syn- corr. to N-terminal telopeptide seq. of hu. type II collagen
US-10-408-178-22
```

```
Query Match          100.0%; Score 26; DB 14; Length 8;
Best Local Similarity 100.0%; Pred. No. 1.2e+06;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 EKAGG 5
    |||||
Db 2 EKAGG 6
```

```
RESULT 13
US-10-408-178-24
; Sequence 24, Application US/10408178
; Publication No. US20030166032A1
; GENERAL INFORMATION:
; APPLICANT: Eyre, David
```

```
; TITLE OF INVENTION: CARTILAGE RESORPTION ASSAYS
; FILE REFERENCE: WROS-1-14269
; CURRENT APPLICATION NUMBER: US/10/408,178
; CURRENT FILING DATE: 2003-04-03
; PRIOR APPLICATION NUMBER: US/10/073,679
; PRIOR FILING DATE: 2002-02-11
; PRIOR APPLICATION NUMBER: 09/385,740
; PRIOR FILING DATE: 1999-08-30
; PRIOR APPLICATION NUMBER: US 60/142,274
; PRIOR FILING DATE: 1999-07-02
; PRIOR APPLICATION NUMBER: US 60/141,574
; PRIOR FILING DATE: 1999-06-29
; PRIOR APPLICATION NUMBER: US 09/335,098
; PRIOR FILING DATE: 1999-06-17
; PRIOR APPLICATION NUMBER: US 60/089,823
; PRIOR FILING DATE: 1998-06-19
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 24
; LENGTH: 8
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: PEPTIDE
; LOCATION: (1)..(8)
; OTHER INFORMATION: syn- corr. to N-terminal telopeptide seq. of hu. type II collagen
US-10-408-178-24
```

```
Query Match          100.0%; Score 26; DB 14; Length 8;
Best Local Similarity 100.0%; Pred. No. 1.2e+06;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 EKAGG 5
    |||||
Db 3 EKAGG 7
```

```
RESULT 14
US-10-615-959-23
; Sequence 23, Application US/10615959
; Publication No. US2004004832A1
; GENERAL INFORMATION:
; APPLICANT: Eyre, David R.
; TITLE OF INVENTION: SYNTHETIC PEPTIDES FOR COLLAGEN RESORPTION ASSAYS
; FILE REFERENCE: WROS-1-18220
; CURRENT APPLICATION NUMBER: US/10/615,959
; CURRENT FILING DATE: 2003-07-08
; PRIOR APPLICATION NUMBER: US/10/009,999A
; PRIOR FILING DATE: 2003-01-22
; PRIOR APPLICATION NUMBER: PCT/US99/29357
; PRIOR FILING DATE: 1999-12-10
; PRIOR APPLICATION NUMBER: US 09/335,098
; PRIOR FILING DATE: 1999-06-17
; PRIOR APPLICATION NUMBER: US 60/141,574
; PRIOR FILING DATE: 1999-06-29
; PRIOR APPLICATION NUMBER: US 60/142,274
; PRIOR FILING DATE: 1999-07-02
; PRIOR APPLICATION NUMBER: US 60/142,675
; PRIOR FILING DATE: 1999-07-07
; PRIOR APPLICATION NUMBER: US 09/385,740
; PRIOR FILING DATE: 1999-08-30
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 23
; LENGTH: 9
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: MISC FEATURE
; OTHER INFORMATION: synthetic peptide corresponding to amino-terminal
; OTHER INFORMATION: telopeptide sequence of human type II collagen
US-10-615-959-23
```

Query Match 100.0%; Score 26; DB 12; Length 9;  
 Best Local Similarity 100.0%; Pred. No. 1.2e+06;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EKAGG 5  
 DB 2 EKAGG 6

RESULT 15  
 US-10-615-959-25  
 ; Sequence 25, Application US/10615959  
 ; Publication No. US20040048321A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Eyre, David R.  
 ; TITLE OF INVENTION: SYNTHETIC PEPTIDES FOR COLLAGEN RESORPTION ASSAYS  
 ; FILE REFERENCE: WROS-1-18220  
 ; CURRENT APPLICATION NUMBER: US/10/615,959  
 ; CURRENT FILING DATE: 2003-07-08  
 ; PRIOR APPLICATION NUMBER: US/10/009,999A  
 ; PRIOR FILING DATE: 2003-01-22  
 ; PRIOR APPLICATION NUMBER: PCT/US99/29357  
 ; PRIOR FILING DATE: 1999-12-10  
 ; PRIOR APPLICATION NUMBER: US 09/335,098  
 ; PRIOR FILING DATE: 1999-06-17  
 ; PRIOR APPLICATION NUMBER: US 60/141,574  
 ; PRIOR FILING DATE: 1999-06-29  
 ; PRIOR APPLICATION NUMBER: US 60/142,274  
 ; PRIOR FILING DATE: 1999-07-02  
 ; PRIOR APPLICATION NUMBER: US 60/142,675  
 ; PRIOR FILING DATE: 1999-07-07  
 ; PRIOR APPLICATION NUMBER: US 09/385,740  
 ; PRIOR FILING DATE: 1999-08-30  
 ; NUMBER OF SEQ ID NOS: 45  
 ; SOFTWARE: PatentIn version 3.2  
 ; SEQ ID NO 25  
 ; LENGTH: 9  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 ; FEATURE:  
 ; NAME/KEY: MISC FEATURE  
 ; OTHER INFORMATION: synthetic peptide corresponding to amino-terminal  
 ; OTHER INFORMATION: telopeptide sequence of human type II collagen  
 US-10-615-959-25

Query Match 100.0%; Score 26; DB 12; Length 9;  
 Best Local Similarity 100.0%; Pred. No. 1.2e+06;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EKAGG 5  
 DB 3 EKAGG 7

Search completed: September 18, 2004, 04:56:16  
 Job time : 25.7255 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 18, 2004, 04:02:05 ; Search time 6.07843 Seconds  
(without alignments)  
79.125 Million cell updates/sec

Title: US-10-615-959-44

Perfect score: 26

Sequence: 1 EKAGG 5

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283366 segs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR\_78:\*  
1: pir1.\*  
2: pir2.\*  
3: pir3.\*  
4: pir4.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	26	100.0	97	B71719	ribosomal protein
2	26	100.0	97	H97716	50S ribosomal prot
3	26	100.0	120	AH1056	hypothetical prote
4	26	100.0	126	A11951	hypothetical prote
5	26	100.0	130	T31248	hypothetical prote
6	26	100.0	141	S18397	hemoglobin alpha c
7	26	100.0	145	B90439	hypothetical prote
8	26	100.0	146	A83014	conserved hypothe
9	26	100.0	146	E98270	hypothetical prote
10	26	100.0	147	S22357	hemoglobin beta-y
11	26	100.0	149	H70441	ribosomal protein
12	26	100.0	153	C97823	50S ribosomal prot
13	26	100.0	156	A83349	LSU ribosomal prot
14	26	100.0	159	G95343	hypothetical prote
15	26	100.0	162	D38257	glucan endo-1,3-be
16	26	100.0	162	E38257	glucan endo-1,3-be
17	26	100.0	164	C81886	hypothetical prote
18	26	100.0	165	B81135	E16-related protei
19	26	100.0	181	D85636	hypothetical prote
20	26	100.0	198	B84842	trp repressor-bind
21	26	100.0	201	D88473	protein F40H6.3 [i
22	26	100.0	202	A45029	brefeldin A resist
23	26	100.0	205	S39863	late competence op
24	26	100.0	206	S18250	collagen alpha 1(I
25	26	100.0	210	D87724	protein F37F2.1 [i
26	26	100.0	210	T33051	hypothetical prote
27	26	100.0	210	E83816	late competence op
28	26	100.0	216	F71002	hypothetical prote
29	26	100.0	216	G75128	hypothetical prote

30	26	100.0	229	2	AE1300	ribonuclease III h
31	26	100.0	229	2	AE1672	ribonuclease III h
32	26	100.0	233	1	S24642	tumor necrosis fac
33	26	100.0	233	2	AB2775	phosphoglycolate p
34	26	100.0	252	2	D69546	extragenic suppres
35	26	100.0	265	2	H97554	phosphoglycolate p
36	26	100.0	278	2	T45802	glucan endo-1,3-be
37	26	100.0	280	2	H81356	probable fructose-
38	26	100.0	281	1	F64306	hypothetical prote
39	26	100.0	293	2	T19954	hypothetical prote
40	26	100.0	298	2	C38257	glucan endo-1,3-be
41	26	100.0	299	2	S20830	polymerase-associ
42	26	100.0	304	2	G86143	probable zinc fing
43	26	100.0	305	2	H75091	hypothetical prote
44	26	100.0	317	2	JC1439	glucan endo-1,3-be
45	26	100.0	318	2	F72255	conserved hypothet

ALIGNMENTS

RESULT 1

B71719  
ribosomal protein L28 - Rickettsia prowazekii  
C:Species: Rickettsia prowazekii  
C:Date: 21-Nov-1998 #sequence\_revision 21-Nov-1998 #text\_change 03-Nov-2000  
C:Accession: B71719  
R:Andersson, S.G.E.; Zomorodipour, A.; Andersson, J.O.; Sichteritz-Ponten, T.; Alsmark, U.  
Nature 396, 133-140, 1998  
A:Title: The genome sequence of Rickettsia prowazekii and the origin of mitochondria.  
A:Reference number: A71630; MUID:99039499; PMID:9823893  
A:Accession: B71719  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-97 <AND>  
A:Cross-references: GB:AJ235270; GB:AJ235269; NID:G3860572; PIDN:CAAL4569.1; PID:el342412  
A:Experimental source: strain Madrid E  
C:Genetics:  
A:Gene: rpmB; RP099

Query Match 100.0%; Score 26; DB 2; Length 97;  
Best Local Similarity 100.0%; Pred. No. 87;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 EKAGG 5  
|||||  
Db 60 EKAGG 64

RESULT 2

H97716  
50S ribosomal protein L28 [imported] - Rickettsia conorii (strain Malish 7)  
C:Species: Rickettsia conorii  
C:Date: 30-Sep-2001 #sequence\_revision 30-Sep-2001 #text\_change 30-Sep-2001  
C:Accession: H97716  
R:Ogata, H.; Audic, S.; Renesto-Audiffren, P.; Fournier, P.E.; Barbe, V.; Samson, D.; Ro  
Science 293, 2093-2098, 2001  
A:Title: Mechanisms of Evolution in Rickettsia conorii and Rickettsia prowazekii.  
A:Reference number: A97700; MUID:21442074; PMID:11557893  
A:Accession: H97716  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-97 <KUR>  
A:Cross-references: GB:AE006914; PIDN:AAL02674.1; PID:g15619179; GSPDB:GN00173  
C:Genetics:  
A:Gene: rpmB

Query Match 100.0%; Score 26; DB 2; Length 97;  
Best Local Similarity 100.0%; Pred. No. 87;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 EKAGG 5  
|||||

Db 60 EKAGG 64

RESULT 3  
AH1056  
hypothetical protein STY4780 [imported] - Salmonella enterica subsp. enterica serovar Typhimurium  
C:Species: Salmonella enterica subsp. enterica serovar Typhi  
A:Note: this species has also been called Salmonella typhi  
C>Date: 09-Nov-2001 #sequence\_revision 09-Nov-2001 #text\_change 18-Nov-2002  
C:Accession: AH1056  
R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher, C.;...  
A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serovar...  
A:Reference number: AB0502; MUID:21534947; PMID:11677608  
A:Accession: AH1056  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-120 <PAR>  
A:Cross-references: GB:AL513382; PIDN:CAD06901.1; PID:gl6505549; GSPDB:GN00176  
C:Genetics:  
A:Gene: STY4780

Query Match 100.0%; Score 26; DB 2; Length 120;  
Best Local Similarity 100.0%; Pred. No. 1e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EKAGG 5  
Db 23 EKAGG 27

RESULT 4  
AH1951  
hypothetical protein all1164 [imported] - Nostoc sp. (strain PCC 7120)  
C:Species: Nostoc sp. PCC 7120  
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120  
C>Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 09-Dec-2002  
C:Accession: AH1951  
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi, N.;...  
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena...  
A:Reference number: AB1807; MUID:21595285; PMID:11759840  
A:Accession: AH1951  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-126 <KUR>  
A:Cross-references: GB:BA000019; PIDN:BA073121.1; PID:gl7130510; GSPDB:GN00179  
A:Experimental source: strain PCC 7120  
C:Genetics:  
A:Gene: all1164

Query Match 100.0%; Score 26; DB 2; Length 126;  
Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EKAGG 5  
Db 24 EKAGG 28

RESULT 5  
T31248  
hypothetical protein 942 - Sphingomonas aromaticivorans plasmid pNL1  
C:Species: Sphingomonas aromaticivorans  
C>Date: 11-Jan-2000 #sequence\_revision 11-Jan-2000 #text\_change 18-Feb-2000  
R:Romine, M.F.; Stillwell, L.C.; Wong, K.K.; Thurston, S.J.; Sisk, E.C.; Senses, C.W.; G...  
A:Title: Complete sequence of a 184 kb catabolic plasmid from Sphingomonas aromaticivorans...  
A:Reference number: AB0502; MUID:21534947; PMID:11677608  
A:Accession: AH1056  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-120 <PAR>  
A:Cross-references: GB:AL513382; PIDN:CAD06901.1; PID:gl6505549; GSPDB:GN00176  
C:Genetics:  
A:Gene: STY4780

Query Match 100.0%; Score 26; DB 2; Length 120;  
Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EKAGG 5  
Db 24 EKAGG 28

RESULT 6  
S18397  
hemoglobin alpha chain - small Madagascar hedgehog  
C:Species: Echinops telfairi (small Madagascar hedgehog)  
C>Date: 19-Mar-1997 #sequence\_revision 19-Mar-1997 #text\_change 19-May-2000  
C:Accession: S18397  
R:Piccinini, M.; Kleinschmidt, T.; Gorr, T.; Weber, R.E.; Kuenzle, H.; Braunitzer, G.  
A:Title: Primary structure and oxygen-binding properties of the hemoglobin from the lesser...  
A:Reference number: S18397; MUID:92172283; PMID:1793518  
A:Accession: S18397  
A:Status: preliminary  
A:Molecule type: protein  
A:Residues: 1-141 <PIC>  
A:Superfamily: globin; globin homology  
C:Keywords: blood; chromoprotein; heme; iron; metalloprotein; oxygen carrier  
F:2-141/Domain: globin homology <GLB>  
F:87/Binding site: heme iron (His) (proximal axial ligand) #status predicted

Query Match 100.0%; Score 26; DB 2; Length 141;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EKAGG 5  
Db 15 EKAGG 19

RESULT 7  
B90439  
hypothetical protein SSO2648 [imported] - Sulfolobus solfataricus  
C:Species: Sulfolobus solfataricus  
C>Date: 24-May-2001 #sequence\_revision 24-May-2001 #text\_change 24-May-2001  
C:Accession: B90439  
R:She, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Awayez, M.J.; Chan-V...  
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena...  
A:Reference number: AB1807; MUID:21595285; PMID:11759840  
A:Accession: AH1951  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-126 <KUR>  
A:Cross-references: GB:BA000019; PIDN:BA073121.1; PID:gl7130510; GSPDB:GN00179  
A:Experimental source: strain PCC 7120  
C:Genetics:  
A:Gene: all1164

Query Match 100.0%; Score 26; DB 2; Length 126;  
Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EKAGG 5  
Db 24 EKAGG 28

RESULT 8  
T31248  
hypothetical protein 942 - Sphingomonas aromaticivorans plasmid pNL1  
C:Species: Sphingomonas aromaticivorans  
C>Date: 11-Jan-2000 #sequence\_revision 11-Jan-2000 #text\_change 18-Feb-2000  
R:Romine, M.F.; Stillwell, L.C.; Wong, K.K.; Thurston, S.J.; Sisk, E.C.; Senses, C.W.; G...  
A:Title: Complete sequence of a 184 kb catabolic plasmid from Sphingomonas aromaticivorans...  
A:Reference number: AB0502; MUID:21534947; PMID:11677608  
A:Accession: AH1056  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-120 <PAR>  
A:Cross-references: GB:AL513382; PIDN:CAD06901.1; PID:gl6505549; GSPDB:GN00176  
C:Genetics:  
A:Gene: STY4780

Query Match 100.0%; Score 26; DB 2; Length 120;  
Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EKAGG 5  
Db 24 EKAGG 28

A:Reference number: Z0992

A:Accession: T31248

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-130 &lt;ROM&gt;

A:Cross-references: EMBL:AF079317; NID:g3378261; PID:g3378389; PIDN:AA03972.1

C:Genetics:

A:Genome: plasmid pNL1

A:Note: orf942

C:Superfamily: hypothetical protein HI0719

Query Match 100.0%; Score 26; DB 2; Length 130;

Best Local Similarity 100.0%; Pred. No. 1.1e+02;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EKAGG 5

Db 62 EKAGG 66

RESULT 6

S18397

hemoglobin alpha chain - small Madagascar hedgehog

C:Species: Echinops telfairi (small Madagascar hedgehog)

C&gt;Date: 19-Mar-1997 #sequence\_revision 19-Mar-1997 #text\_change 19-May-2000

C:Accession: S18397

R:Piccinini, M.; Kleinschmidt, T.; Gorr, T.; Weber, R.E.; Kuenzle, H.; Braunitzer, G.

A:Title: Primary structure and oxygen-binding properties of the hemoglobin from the lesser

A:Reference number: S18397; MUID:92172283; PMID:1793518

A:Accession: S18397

A:Status: preliminary

A:Molecule type: protein

A:Residues: 1-141 &lt;PIC&gt;

A:Superfamily: globin; globin homology

C:Keywords: blood; chromoprotein; heme; iron; metalloprotein; oxygen carrier

F:2-141/Domain: globin homology &lt;GLB&gt;

F:87/Binding site: heme iron (His) (proximal axial ligand) #status predicted

Query Match

Best Local Similarity 100.0%; Score 26; DB 2; Length 141;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EKAGG 5

Db 15 EKAGG 19

RESULT 7

B90439

hypothetical protein SSO2648 [imported] - Sulfolobus solfataricus

C:Species: Sulfolobus solfataricus

C&gt;Date: 24-May-2001 #sequence\_revision 24-May-2001 #text\_change 24-May-2001

C:Accession: B90439

R:She, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Awayez, M.J.; Chan-V

A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena

A:Reference number: AB1807; MUID:21595285; PMID:11759840

A:Accession: AH1951

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-145 &lt;KUR&gt;

A:Cross-references: GB:AE006641; NID:gl3815963; PIDN:AAK42769.1; GSPDB:GN00155

C:Genetics:

A:Gene: SSO2648

Query Match 100.0%; Score 26; DB 2; Length 145;

Best Local Similarity 100.0%; Pred. No. 1.2e+02;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EKAGG 5

Db 57 EKAGG 61  
|||||

## RESULT 8

AB3014  
conserved hypothetical protein Atu3718 [imported] - Agrobacterium tumefaciens (strain C58)  
C/Species: Agrobacterium tumefaciens  
C/Date: 11-Jan-2002 #sequence\_revision 11-Jan-2002 #text\_change 18-Nov-2002  
C/Accession: AB3014  
R/Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I.; Erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; McClellan, Karp, P.; Romero, P.; Zhang, S.  
Science 294, 2317-2323, 2001  
A/Authors: Yoo, H.; Iao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm, ster, E.W.  
A/Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.  
A/Reference number: AB2577; MUID:21608550; PMID:11743193  
A/Accession: AB3014  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-146 <KUR>  
A/Cross-references: GB:AE008689; PIDN:AAL44528.1; PID:gi7742141; GSPDB:GN00187  
A/Experimental source: strain C58 (Dupont)  
C/Genetics:  
A/Gene: Atu3718  
A/Map position: linear chromosome  
C/Superfamily: 15.5K protein (tolAB operon 5' region)

Query Match 100.0%; Score 26; DB 2; Length 146;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 EKAGG 5  
|||||  
Db 97 EKAGG 101

## RESULT 9

E98270  
hypothetical protein AGR\_L\_2234 [imported] - Agrobacterium tumefaciens (strain C58, Cere)  
C/Species: Agrobacterium tumefaciens  
C/Date: 22-Oct-2001 #sequence\_revision 22-Oct-2001 #text\_change 18-Nov-2002  
C/Accession: E98270  
R/Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Quorollo, B.; Goldman, A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.; Science 294, 2323-2328, 2001  
A/Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tumefaciens  
A/Reference number: A97359; MUID:21608551; PMID:11743194  
A/Accession: E98270  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-146 <KUR>  
A/Cross-references: GB:AE007870; PIDN:AAK89687.1; PID:g15159593; GSPDB:GN00170  
C/Genetics:  
A/Gene: AGR\_L\_2234  
A/Map position: linear chromosome  
C/Superfamily: 15.5K protein (tolAB operon 5' region)

Query Match 100.0%; Score 26; DB 2; Length 146;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 EKAGG 5  
|||||  
Db 97 EKAGG 101

## RESULT 10

S22357  
hemoglobin beta-y chain, embryonic - golden hamster  
C/Species: Mesocricetus auratus (golden hamster)  
C/Date: 22-Nov-1993 #sequence\_revision 01-Dec-1995 #text\_change 04-Mar-2000

C/Accession: S22357

R/Li, H.; Subar, M.; Lee, K.M.; Boussios, T.  
Biochim. Biophys. Acta 1130, 218-220, 1992  
A/Title: Cloning and sequence analysis of two embryonic beta-like globin cDNAs (y and z)  
A/Reference number: S22356; MUID:92223098; PMID:13339291  
A/Accession: S22357

A/Status: preliminary

A/Molecule type: mRNA

A/Residues: 1-147 <LIH>

A/Cross-references: EMBL:X64179; NID:g49617; PIDN:CAA45518.1; PID:g49618

C/Superfamily: globin; globin homology

C/Keywords: Chromoprotein; heme; iron; metalloprotein; oxygen carrier

F/4-147/Domain: globin homology <GUB>

F/64/Binding site: oxygen (His) (distal axial ligand) #status predicted

F/93/Binding site: heme iron (His) (proximal axial ligand) #status predicted

Query Match 100.0%; Score 26; DB 2; Length 147;  
Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 EKAGG 5  
|||||

Db 22 EKAGG 26

## RESULT 11

H70441

ribosomal protein L15 - Aquifex aeolicus

C/Species: Aquifex aeolicus

C/Date: 08-May-1998 #sequence\_revision 08-May-1998 #text\_change 13-Aug-1999

C/Accession: H70441

R/Deckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; Ove

V. Nature 392, 353-358, 1998

A/Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.  
A/Reference number: A70300; MUID:98196666; PMID:9537320

A/Accession: H70441

A/Status: preliminary; nucleic acid sequence not shown; translation not shown

A/Molecule type: DNA

A/Residues: 1-149 <AQF>

A/Cross-references: GB:AE000749; NID:g2983975; PIDN:AAC07532.1; PID:g2983986; GB:AE00065;

A/Experimental source: strain VF5

C/Genetics:

A/Gene: rplO

C/Superfamily: Escherichia coli ribosomal protein L15

Query Match 100.0%; Score 26; DB 2; Length 149;  
Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 EKAGG 5  
|||||

Db 139 EKAGG 143

## RESULT 12

C97823

50S ribosomal protein L15 [imported] - Rickettsia conorii (strain Malish 7)

C/Species: Rickettsia conorii

C/Date: 30-Sep-2001 #sequence\_revision 30-Sep-2001 #text\_change 22-Oct-2001

C/Accession: C97823

R/Ogata, H.; Audic, S.; Renesto-Audiffren, P.; Fournier, P.E.; Barbe, V.; Samson, D.; Ro

science 293, 2093-2098, 2001

A/Title: Mechanisms of Evolution in Rickettsia conorii and Rickettsia prowazekii.

A/Reference number: A97700; MUID:21442074; PMID:11557893

A/Accession: C97823

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-153 <KUR>

A/Cross-references: GB:AE006914; PIDN:AAL03525.1; PID:g15620101; GSPDB:GN00173

C/Genetics:

A/Gene: rplO

C/Superfamily: Escherichia coli ribosomal protein L15

Query Match 100.0%; Score 26; DB 2; Length 153;  
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EXAGG 5  
 |||||  
 Db 146 EXAGG 150

## RESULT 13

A33349  
 LSU ribosomal protein L15P [imported] - Brucella melitensis (strain 16M)  
 C:Species: Brucella melitensis  
 C:Date: 01-Feb-2002 #sequence\_revision 01-Feb-2002 #text\_change 15-Feb-2002  
 C:Accession: AB3349  
 R:DelVecchio, V.G.; Kapatal, V.; Redkar, R.J.; Patra, G.; Mujer, C.; Los, T.; Ivanova,  
 ; Mazur, M.; Goltzman, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Letess  
 Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002  
 A:Title: The genome sequence of the facultative intracellular pathogen Brucella melitensis  
 A:Reference number: AD3252; PMID:11756688  
 A:Accession: AB3349  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-156 <XUR>  
 A:Cross-references: GB:AE008917; PIDN:AAL51957.1; PID:g17982716; GSPDB:GNC0190  
 A:Experimental source: strain 16M  
 C:Genetics:  
 A:Gene: BMEI0776  
 A:Map position: 1  
 C:Superfamily: Escherichia coli ribosomal protein L15

Query Match 100.0%; Score 26; DB 2; Length 156;  
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EXAGG 5  
 |||||  
 Db 142 EXAGG 146

## RESULT 14

G95343  
 Hypothetical protein Smal200 [imported] - Sinorhizobium meliloti (strain 1021) magaplaasm  
 C:Species: Sinorhizobium meliloti  
 C:Date: 24-Aug-2001 #sequence\_revision 24-Aug-2001 #text\_change 30-Sep-2001  
 C:Accession: G95343  
 R:Barnett, M.J.; Fisher, R.F.; Jones, T.; Komp, C.; Abola, A.P.; Barloy-Hubler, F.; Bows  
 ; Kalman, S.; Keating, D.H.; Palm, C.; Peck, M.C.; Surzycki, R.; Wells, D.H.; Yeh, K.C.  
 Proc. Natl. Acad. Sci. U.S.A. 98, 9893-9898, 2001  
 A:Title: Nucleotide sequence and predicted functions of the entire Sinorhizobium meliloti  
 A:Reference number: A95262; MUID:21396509; PMID:11481432  
 A:Accession: G95343  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-159 <XUR>  
 A:Cross-references: GB:AE006469; PIDN:AAK65313.1; PID:g14523769; GSPDB:GNC00165  
 A:Experimental source: strain 1021, megaplaasmid pSymA  
 R:Galibert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubler,  
 pela, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.;  
 L.; Hyman, R.W.; Jones, T.  
 Science 293, 668-672, 2001  
 A:Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelaure,  
 hebaud, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K.  
 A:Title: The composite genome of the legume symbiont Sinorhizobium meliloti.  
 A:Reference number: A96039; MUID:21368234; PMID:11474104  
 A:Contents: annotation  
 C:Genetics:  
 A:Gene: Smal200  
 A:Genome: plasmid

Query Match 100.0%; Score 26; DB 2; Length 159;  
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 EXAGG 5  
 |||||  
 Db 154 EXAGG 158

## RESULT 15

D38257  
 glucan endo-1,3-beta-D-glucosidase (EC 3.2.1.39), acidic (clone ci30) - common tobacco (c  
 C:Species: Nicotiana tabacum (common tobacco)  
 C:Date: 14-Jun-1991 #sequence\_revision 14-Jun-1991 #text\_change 13-Sep-1998  
 C:Accession: D38257  
 R:Linthorst, H.J.M.; Melchers, L.S.; Mayer, A.; Van Roekel, J.S.C.; Cornelissen, B.J.C.;  
 Proc. Natl. Acad. Sci. U.S.A. 87, 8756-8760, 1990  
 A:Title: Analysis of gene families encoding acidic and basic beta-1,3-glucanases of tobac  
 A:Reference number: A38257; MUID:91062359; PMID:2247445  
 A:Accession: D38257  
 A:Status: preliminary; not compared with conceptual translation  
 A:Molecule type: mRNA  
 A:Residues: 1-162 <LIN>  
 C:Superfamily: beta-1,3-glucanase  
 C:Keywords: glycosidase; hydrolase; polysaccharide degradation

Query Match 100.0%; Score 26; DB 2; Length 162;  
 Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EXAGG 5  
 |||||  
 Db 70 EXAGG 74

Search completed: September 18, 2004, 04:30:56

Job time : 8.07843 secs

GenCore version 5.1.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 18, 2004, 03:48:54 ; Search time 3.72549 Seconds  
(without alignments)  
69.884 Million cell updates/sec

Title: US-10-615-959-44

Perfect score: 26

Sequence: 1 EKAGG 5

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues  
Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_42:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query %	Length	DB ID	Description
1	26	100.0	97	1 RL28_RICCN	Q92jdl rickettsia
2	26	100.0	97	1 RL28_RICPR	Q9ze48 rickettsia
3	26	100.0	141	1 HBA_ECHTE	P24291 echinops te
4	26	100.0	146	1 HBBY_MESAU	P29626 mesocricetu
5	26	100.0	149	1 RL15_AQUAE	O67561 aquifex aeo
6	26	100.0	160	1 RL13_TOBAC	P23397 nicotiana t
7	26	100.0	180	1 RL15_LEPIN	Q9xd17 leptospira
8	26	100.0	187	1 RL18_ARATH	P42791 arabidopsis
9	26	100.0	193	1 APT_BIFLO	Q8g6b5 bifidobacte
10	26	100.0	197	1 WRBA_ECO57	O8x4b4 escherichia
11	26	100.0	197	1 WRBA_ECOLI	P30849 escherichia
12	26	100.0	198	1 HB2G_HUMAN	P01911 homo sapien
13	26	100.0	202	1 F25_SCHPO	P30821 schizosacch
14	26	100.0	205	1 CMEI_BACSU	P39694 bacillus su
15	26	100.0	229	1 RNC_LISIN	Q92ak3 listeria su
16	26	100.0	229	1 RNC_LISMO	O8y691 listeria mo
17	26	100.0	233	1 GPH_AGR75	Q8uey9 agrobacteri
18	26	100.0	233	1 TNFA_BOVIN	Q06599 bos taurus
19	26	100.0	233	1 TNFA_BUBBU	P59693 bubalus bub
20	26	100.0	234	1 TNFA_BOVIN	P59694 bos indicus
21	26	100.0	252	1 SUHE_ARCFU	O30298 archaeoglob
22	26	100.0	275	1 E131_TOBAC	P52396 nicotiana t
23	26	100.0	281	1 Y054_METJA	Q60361 methanococc
24	26	100.0	287	1 EFTS_PSEPK	Q88mh9 pseudomonas
25	26	100.0	299	1 VV_MEASY	P60168 nealess vir
26	26	100.0	312	1 YD63_THEAC	Q9hih6 thermoplasm
27	26	100.0	321	1 E13F_HORVU	O02439 hordeum vul
28	26	100.0	322	1 RM10_YEAST	P36520 saccharomyc
29	26	100.0	331	1 E13K_TOBAC	P52398 nicotiana t
30	26	100.0	333	1 RBKS_YEAST	P25332 saccharomyc
31	26	100.0	336	1 F16P_SHEEP	P09199 oviss aries
32	26	100.0	337	1 F16P_HUMAN	P09467 homo sapien
33	26	100.0	337	1 F16P_MOUSE	Q9qxd6 mus musculu

34 26 100.0 337 1 F16P\_PIG  
35 26 100.0 337 1 F16P\_RABIT  
36 26 100.0 343 1 E13G\_TOBAC  
37 26 100.0 347 1 E13A\_SOYBN  
38 26 100.0 350 1 E13B\_PRUPE  
39 26 100.0 351 1 E13C\_TOBAC  
40 26 100.0 351 1 E13D\_TOBAC  
41 26 100.0 356 1 E13L\_TOBAC  
42 26 100.0 362 1 F16P\_RAT  
43 26 100.0 371 1 H2AW\_HUMAN  
44 26 100.0 375 1 TGT\_THETN  
45 26 100.0 387 1 S17P\_SPIOL

P00636 sus scrofa  
P00637 oryctolagus  
P23547 nicotiana t  
Q03773 glycine max  
P52408 prunus pers  
P23432 nicotiana t  
P23433 nicotiana t  
P52399 nicotiana t  
P19112 rattus norv  
Q9p0m6 homo sapien  
Q8ram9 thermoaer  
O20252 spinacia ol

## ALIGNMENTS

RESULT 1  
RL28\_RICCN  
ID RL28\_RICCN STANDARD; PRT; 97 AA.  
AC Q92jdl;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DE 28-FEB-2003 (Rel. 41, Last annotation update)  
DE 50S ribosomal protein L28.  
GN RPMB OR RC0136.  
OS Rickettsia conorii.  
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;  
OC Rickettsiaceae; Rickettsiae; Rickettsia.  
OX NCBI\_TaxID=781;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Malish 7;  
RX MEDLINE=21442074; PubMed=11557893;  
RA Ogata H., Audic S., Renesto-Audiffren P., Fournier P.-E., Barbe V.,  
RA Samson D., Roux V., Cossart P., Weissenbach J., Claverie J.-M.,  
RA Raoult D.;  
RT "Mechanisms of evolution in Rickettsia conorii and R. prowazekii.";  
RL Science 293:2093-2098(2001).  
CC -!- SIMILARITY: Belongs to the L28P family of ribosomal proteins.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; AE008580; AA02674.1; -.  
DR PIR; H97716; H97716.  
DR HAMAP; MF 00373; -; 1.  
DR InterPro; IPR001383; Ribosomal L28.  
DR Pfam; PF00830; Ribosomal\_L28; 1.  
DR TIGRfams; TIGR00009; L28; 1.  
KW Ribosomal protein; Complete proteome.  
SQ SEQUENCE 97 AA; 10766 MW; DFF539E8531596C CRC64;  
  
Query Match 100.0%; Score 26; DB 1; Length 97;  
Best Local Similarity 100.0%; Pred. No. 34;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 EKAGG 5  
Db 60 EKAGG 64  
|||||  
  
RESULT 2  
RL28\_RICPR  
ID RL28\_RICPR STANDARD; PRT; 97 AA.  
AC Q9ZE48;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)

DT	28-FEB-2003 (Rel. 41, Last annotation update)	CC	!!- TISSUE SPECIFICITY: Red blood cells.
DE	50S ribosomal protein L28.	CC	!!- SIMILARITY: Belongs to the globin family.
PM	OR RP099.	DR	PIR; S18397; S18397.
OS	Rickettsia prowazekii.	DR	HSP; P01958; 2MRB.
OC	Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;	DR	InterPro; IPR002338; Alpha haem.
CC	Rickettsiaceae; Rickettsiinae; Rickettsia.	DR	InterPro; IPR000971; Globin.
OX	NCBI_TaxID=782;	DR	Pfam; PF00042; Globin; 1.
RN	[1]	DR	PRINTS; PR00612; ALPHAHAE.
RP	SEQUENCE FROM N.A.	DR	PROSITE; PS01033; GLOBIN; 1.
RC	STRAIN=Madrid E;	KW	Heme; Oxygen transport; Transport; Erythrocyte.
RX	MEDLINE=99039499; PubMed=9823893;	FT	IRON (HEME DISTAL LIGAND).
RA	Anderson S.G.E., Zomorodipour A., Andersson J.O.,	FT	IRON (HEME PROXIMAL LIGAND).
RA	Sicheritz-Ponten T., Almaraz U.C.M., Podowski R.M., Naeslund A.K.,	FT	IRON (HEME PROXIMAL LIGAND).
RA	Eriksson A.-S., Winkler H.H., Kurland C.G.;	FT	IRON (HEME PROXIMAL LIGAND).
RT	"The genome sequence of Rickettsia prowazekii and the origin of	FT	IRON (HEME PROXIMAL LIGAND).
RT	mitochondria."	FT	IRON (HEME PROXIMAL LIGAND).
RL	Nature 396:133-140(1998).	FT	IRON (HEME PROXIMAL LIGAND).
CC	!!- SIMILARITY: Belongs to the L28P family of ribosomal proteins.	FT	IRON (HEME PROXIMAL LIGAND).
CC	-----	FT	IRON (HEME PROXIMAL LIGAND).
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration	FT	IRON (HEME PROXIMAL LIGAND).
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -	FT	IRON (HEME PROXIMAL LIGAND).
CC	the European Bioinformatics Institute. There are no restrictions on its	FT	IRON (HEME PROXIMAL LIGAND).
CC	use by non-profit institutions as long as its content is in no way	FT	IRON (HEME PROXIMAL LIGAND).
CC	modified and this statement is not removed. Usage by and for commercial	FT	IRON (HEME PROXIMAL LIGAND).
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>	FT	IRON (HEME PROXIMAL LIGAND).
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).	FT	IRON (HEME PROXIMAL LIGAND).
CC	-----	FT	IRON (HEME PROXIMAL LIGAND).
CC	EMBL; AJ235270; CAA14569.1; -.	FT	IRON (HEME PROXIMAL LIGAND).
DR	PIR; B71719; B71719.	FT	IRON (HEME PROXIMAL LIGAND).
DR	HMAP; MF 00373; -; 1.	FT	IRON (HEME PROXIMAL LIGAND).
DR	InterPro; IPR001383; Ribosomal L28.	FT	IRON (HEME PROXIMAL LIGAND).
DR	Pfam; PF00830; Ribosomal L28; 1.	FT	IRON (HEME PROXIMAL LIGAND).
DR	TIGRFAM; TIGR00009; L28; 1.	FT	IRON (HEME PROXIMAL LIGAND).
KW	Ribosomal protein; Complete proteome.	FT	IRON (HEME PROXIMAL LIGAND).
SQ	SEQUENCE 97 AA; 10953 MW; 96889B2FAE1E28C8 CRC64;	FT	IRON (HEME PROXIMAL LIGAND).
Query Match	100.0%; Score 26; DB 1; Length 97;	FT	IRON (HEME PROXIMAL LIGAND).
Best Local Similarity	100.0%; Pred. No. 34;	FT	IRON (HEME PROXIMAL LIGAND).
Matches	5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	FT	IRON (HEME PROXIMAL LIGAND).
QY	1 EXAGG 5	FT	IRON (HEME PROXIMAL LIGAND).
Db	60 EXAGG 64	FT	IRON (HEME PROXIMAL LIGAND).
RESULT 3		FT	IRON (HEME PROXIMAL LIGAND).
HBA_ECHTE		FT	IRON (HEME PROXIMAL LIGAND).
ID	HBA_ECHTE STANDARD; PRT; 141 AA.	FT	IRON (HEME PROXIMAL LIGAND).
AC	P24231; (Rel. 21, Created)	FT	IRON (HEME PROXIMAL LIGAND).
DT	01-MAR-1992 (Rel. 21, Last sequence update)	FT	IRON (HEME PROXIMAL LIGAND).
DT	01-MAR-1992 (Rel. 21, Last sequence update)	FT	IRON (HEME PROXIMAL LIGAND).
DT	28-FEB-2003 (Rel. 41, Last annotation update)	FT	IRON (HEME PROXIMAL LIGAND).
DE	Hemoglobin alpha chain.	FT	IRON (HEME PROXIMAL LIGAND).
GN	HBA.	FT	IRON (HEME PROXIMAL LIGAND).
OS	Echinops telfairi (Lesser hedgehog tenrec).	FT	IRON (HEME PROXIMAL LIGAND).
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;	FT	IRON (HEME PROXIMAL LIGAND).
OC	Mammalia; Eutheria; Insectivora; Tenrecidae; Tenrecinae; Echinops.	FT	IRON (HEME PROXIMAL LIGAND).
OX	NCBI_TaxID=9371;	FT	IRON (HEME PROXIMAL LIGAND).
RN	[1]	FT	IRON (HEME PROXIMAL LIGAND).
RP	SEQUENCE	FT	IRON (HEME PROXIMAL LIGAND).
RX	MEDLINE=92172283; PubMed=1793518;	FT	IRON (HEME PROXIMAL LIGAND).
RA	Piccinini M., Kleinschmidt T., Gorr T., Weber R.E., Kuenzle H.,	FT	IRON (HEME PROXIMAL LIGAND).
RA	Braunitzer G.;	FT	IRON (HEME PROXIMAL LIGAND).
RT	"Primary structure and oxygen-binding properties of the hemoglobin	FT	IRON (HEME PROXIMAL LIGAND).
RT	from the lesser hedgehog tenrec (Echinops telfairi, Zalambodonta).	FT	IRON (HEME PROXIMAL LIGAND).
RT	Evidence for phylogenetic isolation."	FT	IRON (HEME PROXIMAL LIGAND).
EL	Biol. Chem. Hoppe-Seyler 372:975-989(1991).	FT	IRON (HEME PROXIMAL LIGAND).
CC	!!- FUNCTION: Involved in oxygen transport from the lung to the	FT	IRON (HEME PROXIMAL LIGAND).
CC	various peripheral tissues.	FT	IRON (HEME PROXIMAL LIGAND).
CC	!!- SUBUNIT: Heterotetramer of two alpha chains and two beta chains.	FT	IRON (HEME PROXIMAL LIGAND).
CC	Two external cysteine residues at beta-16 and beta-52 cause	FT	IRON (HEME PROXIMAL LIGAND).
CC	reversible polymerization to octamers and most likely	FT	IRON (HEME PROXIMAL LIGAND).
CC	irreversible formation of higher polymers.	FT	IRON (HEME PROXIMAL LIGAND).
CC	-----	FT	IRON (HEME PROXIMAL LIGAND).
CC	EMBL; X64179; CAA45518.1; -.	FT	IRON (HEME PROXIMAL LIGAND).
DR	PIR; S22357; S22357.	FT	IRON (HEME PROXIMAL LIGAND).
DR	HSP; P02100; 1A9W.	FT	IRON (HEME PROXIMAL LIGAND).
DR	InterPro; IPR002337; Beta haem.	FT	IRON (HEME PROXIMAL LIGAND).
DR	InterPro; IPR000971; Globin.	FT	IRON (HEME PROXIMAL LIGAND).
DR	Pfam; PF00042; globin; 1.	FT	IRON (HEME PROXIMAL LIGAND).
DR	PRINTS; PR00814; BETAHAEM.	FT	IRON (HEME PROXIMAL LIGAND).
DR	PROSITE; PS01033; GLOBIN; 1.	FT	IRON (HEME PROXIMAL LIGAND).
KW	Heme; Oxygen transport; Transport; Erythrocyte; Embryo.	FT	IRON (HEME PROXIMAL LIGAND).
FT	INIT MET 0	FT	IRON (HEME PROXIMAL LIGAND).
FT	METAL 63	FT	IRON (HEME PROXIMAL LIGAND).
FT	METAL 92	FT	IRON (HEME PROXIMAL LIGAND).
FT	SEQUENCE 146 AA; 16274 MW; 2E7D0D46898DFF4F CRC64;	FT	IRON (HEME PROXIMAL LIGAND).
SQ		FT	IRON (HEME PROXIMAL LIGAND).
Query Match	100.0%; Score 26; DB 1; Length 146;	FT	IRON (HEME PROXIMAL LIGAND).
Best Local Similarity	100.0%; Pred. No. 50;	FT	IRON (HEME PROXIMAL LIGAND).
Matches	5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	FT	IRON (HEME PROXIMAL LIGAND).
QY	1 EXAGG 5	FT	IRON (HEME PROXIMAL LIGAND).
Db	15 EXAGG 19	FT	IRON (HEME PROXIMAL LIGAND).
RESULT 4		FT	IRON (HEME PROXIMAL LIGAND).
HBBY_MESAU		FT	IRON (HEME PROXIMAL LIGAND).
ID	HBBY_MESAU STANDARD; PRT; 146 AA.	FT	IRON (HEME PROXIMAL LIGAND).
AC	P29676;	FT	IRON (HEME PROXIMAL LIGAND).
DT	01-APR-1993 (Rel. 25, Created)	FT	IRON (HEME PROXIMAL LIGAND).
DT	01-APR-1993 (Rel. 25, Last sequence update)	FT	IRON (HEME PROXIMAL LIGAND).
DT	28-FEB-2003 (Rel. 41, Last annotation update)	FT	IRON (HEME PROXIMAL LIGAND).
DE	Hemoglobin beta-Y chain.	FT	IRON (HEME PROXIMAL LIGAND).
GN	HBBY.	FT	IRON (HEME PROXIMAL LIGAND).
OS	Mesocricetus auratus (Golden hamster).	FT	IRON (HEME PROXIMAL LIGAND).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	FT	IRON (HEME PROXIMAL LIGAND).
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;	FT	IRON (HEME PROXIMAL LIGAND).
OC	Mesocricetus.	FT	IRON (HEME PROXIMAL LIGAND).
OX	NCBI_TaxID=10036;	FT	IRON (HEME PROXIMAL LIGAND).
RN	[1]	FT	IRON (HEME PROXIMAL LIGAND).
RP	SEQUENCE FROM N.A.	FT	IRON (HEME PROXIMAL LIGAND).
RX	MEDLINE=92223098; PubMed=1339291;	FT	IRON (HEME PROXIMAL LIGAND).
RA	Li H., Subar M., Lee K.M., Bousios T.;	FT	IRON (HEME PROXIMAL LIGAND).
RT	"Cloning and sequence analysis of two embryonic beta-like globin	FT	IRON (HEME PROXIMAL LIGAND).
RT	cDNAs (y and z) of hamster."	FT	IRON (HEME PROXIMAL LIGAND).
RL	Biochim. Biophys. Acta 1130:218-220(1992).	FT	IRON (HEME PROXIMAL LIGAND).
CC	!!- FUNCTION: This is a minor early embryonic beta chain.	FT	IRON (HEME PROXIMAL LIGAND).
CC	!!- SUBUNIT: Heterotetramer of two alpha chains and two beta chains.	FT	IRON (HEME PROXIMAL LIGAND).
CC	!!- SIMILARITY: Belongs to the globin family.	FT	IRON (HEME PROXIMAL LIGAND).
CC	-----	FT	IRON (HEME PROXIMAL LIGAND).
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration	FT	IRON (HEME PROXIMAL LIGAND).
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -	FT	IRON (HEME PROXIMAL LIGAND).
CC	the European Bioinformatics Institute. There are no restrictions on its	FT	IRON (HEME PROXIMAL LIGAND).
CC	use by non-profit institutions as long as its content is in no way	FT	IRON (HEME PROXIMAL LIGAND).
CC	modified and this statement is not removed. Usage by and for commercial	FT	IRON (HEME PROXIMAL LIGAND).
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>	FT	IRON (HEME PROXIMAL LIGAND).
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).	FT	IRON (HEME PROXIMAL LIGAND).
CC	-----	FT	IRON (HEME PROXIMAL LIGAND).
CC	EMBL; X64179; CAA45518.1; -.	FT	IRON (HEME PROXIMAL LIGAND).
DR	PIR; S22357; S22357.	FT	IRON (HEME PROXIMAL LIGAND).
DR	HSP; P02100; 1A9W.	FT	IRON (HEME PROXIMAL LIGAND).
DR	InterPro; IPR002337; Beta haem.	FT	IRON (HEME PROXIMAL LIGAND).
DR	InterPro; IPR000971; Globin.	FT	IRON (HEME PROXIMAL LIGAND).
DR	Pfam; PF00042; globin; 1.	FT	IRON (HEME PROXIMAL LIGAND).
DR	PRINTS; PR00814; BETAHAEM.	FT	IRON (HEME PROXIMAL LIGAND).
DR	PROSITE; PS01033; GLOBIN; 1.	FT	IRON (HEME PROXIMAL LIGAND).
KW	Heme; Oxygen transport; Transport; Erythrocyte; Embryo.	FT	IRON (HEME PROXIMAL LIGAND).
FT	INIT MET 0	FT	IRON (HEME PROXIMAL LIGAND).
FT	METAL 63	FT	IRON (HEME PROXIMAL LIGAND).
FT	METAL 92	FT	IRON (HEME PROXIMAL LIGAND).
FT	SEQUENCE 146 AA; 16274 MW; 2E7D0D46898DFF4F CRC64;	FT	IRON (HEME PROXIMAL LIGAND).
SQ		FT	IRON (HEME PROXIMAL LIGAND).
Query Match	100.0%; Score 26; DB 1; Length 146;	FT	IRON (HEME PROXIMAL LIGAND).
Best Local Similarity	100.0%; Pred. No. 50;	FT	IRON (HEME PROXIMAL LIGAND).
Matches	5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	FT	IRON (HEME PROXIMAL LIGAND).
QY	1 EXAGG 5	FT	IRON (HEME PROXIMAL LIGAND).
Db	15 EXAGG 19	FT	IRON (HEME PROXIMAL LIGAND).





```

RP SEQUENCE FROM N.A.
RC STRAIN=Lai / Serogroup Icterohaemorrhagiae / Serovar lai;
RX MEDLINE=2008835; PubMed=10620683;
RA Zuercher R.L., Hartskeel R.A., van de Kemp H., Bal A.E.;
RT "Characterization of the Leptospira interrogans S10-spc-alpha
RL operon.";
RL FEMS Microbiol. Lett. 182:303-308(2000).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=56601 / Serogroup Icterohaemorrhagiae / Serovar lai;
RX MEDLINE=22598143; PubMed=12712204;
RA Ren S.-X., Fu G., Jiang X.-G., Zeng R., Miao Y.-G., Xu H.,
RA Zhang Y.-X., Xiong H., Lu G., Lu L.-F., Jiang H.-Q., Jia J., Tu Y.-F.,
RA Jiang Y.-X., Gu W.-Y., Zhang Y.-Q., Cai Z., Sheng H.-H., Yin H.-F.,
RA Zhang Y., Zhu G.-F., Wan M., Huang H.-L., Qian Z., Wang S.-Y., Ma W.,
RA Yao Z.-J., Shen Y., Qiang B.-Q., Xia Q.-C., Guo X.-K., Danchin A.,
RA Saint Girons I., Somerville R.L., Wen Y.-M., Shi M.-H., Chen Z.,
RA Xu J.-G., Zhao G.-P.;
RT "Unique physiological and pathogenic features of Leptospira
RT interrogans revealed by whole-genome sequencing.";
RL Nature 422:888-893(2003).
RL [2]
CC -!- FUNCTION: This protein binds directly to 23S ribosomal RNA (By
CC similarity).
CC -!- SIMILARITY: Belongs to the L15P family of ribosomal proteins.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
DR EMBL; AF115283; AAD40602.1; -.
DR EMBL; AB011262; AAN47957.1; -.
DR InterPro; IPR001196; Ribosomal_L15.
DR InterPro; IPR005749; Ribosomal_L15_b.
DR Pfam; PF00256; L15; 1.
DR Pfam; PF01305; Ribosomal_L15; 1.
DR TIGRFAM; TIGR01071; rplO_bact; 1.
DR PROSITE; PS00475; RIBOSOMAL_L15; 1.
KW Ribosomal protein; rRNA-binding; Complete proteome.
SQ SEQUENCE 180 AA; 19865 MW; 2B9D951F905A0602 CRC64;

Query Match 100.0%; Score 26; DB 1; Length 180;
Best Local Similarity 100.0%; Pred. No. 61;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 EXAGG 5
Db 159 EXAGG 163
|||||

RESULT 8
RL18 ARATH
ID RL18 ARATH STANDARD; PRT; 187 AA.
AC P42751; Q9M9W4;
DT 01-NOV-1995 (Rel. 32, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE 60S ribosomal protein L18.
GN RPL18 OR AT3G05590 OR F18C1.14.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids II; Brassicales; Brassicaceae; Arabidopses.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=95180713; PubMed=7875584;
RA Balma S., Sessa G., Ruberti I., Morelli G.;
RT "A cDNA encoding Arabidopsis thaliana cytoplasmic ribosomal protein

```

```

RT L18.";
RL Gene 153:171-174(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=21016720; PubMed=11130713;
RA Salanoubat M., Lemcke K., Rieger M., Ansoerge W., Unseld M.,
RA Fartmann B., Valle G., Bloeker H., Perez-Alonso M., Obermaier B.,
RA Delseny M., Boutry M., Grivell L.A., Mache R., Puigdomenech P.,
RA de Simone V., Choisine N., Artiguenave F., Robert C., Brottier P.,
RA Wincker P., Cattolico L., Weissbach J., Saurin W., Quetier F.,
RA Schaefer M., Mueller-Auer S., Gabel C., Fuchs M., Benes V.,
RA Wurmbach E., Drzonek H., Erle H., Jordan N., Bangert S.,
RA Wiedemann R., Kranz H., Voss H., Holland R., Brandt P., Nyakatura G.,
RA Vezzi A., D'Angelo M., Pallavicini A., Toppo S., Simonetti B.,
RA Conrad A., Hornischer K., Kauer G., Loehner T.-H., Nordstiek G.,
RA Reichelt J., Schaefer M., Schoen O., Barges M., Terol J., Clement J.,
RA Navarro P., Collado C., Perez-Perez A., Ottenwaelder B., Duchemin D.,
RA Cooke R., Laudie M., Berger-Llauro C., Purnelle B., Masuy D.,
RA de Haan M., Maarse A.C., Alcaraz J.-P., Cottet A., Casacuberta E.,
RA Monfort A., Argirou A., Flores M., Liguori R., Vitale D.,
RA Mannhaupt G., Haase D., Schoof H., Rued S., Zaccaria P., Mewes H.-W.,
RA Mayer K.F.X., Kaul S., Town C.D., Koo H.L., Tallon L.J., Jenkins J.,
RA Rooney T., Rizzo M., Walts A., Utterback T., Fujii C.Y., Shea T.P.,
RA Creasy T.H., Haas B., Maiti R., Wu D., Peterson J., Van Aken S.,
RA Pai G., Miltscher J., Sellers P., Gill J.E., Feldblyum T.V.,
RA Preuss D., Lin X., Niernan W.C., Salzberg S.L., White O., Venter J.C.,
RA Fraser C.M., Kaneko T., Nakamura Y., Sato S., Kato T., Asamizu E.,
RA Sasaki S., Kimura T., Ideawa K., Kawashima K., Kishida Y.,
RA Kiyokawa C., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
RA Nakayama S., Nakazaki N., Shinpo S., Takeuchi C., Wada T.,
RA Watanabe A., Yamada M., Yasuda M., Tabata S.;
RT "Sequence and analysis of chromosome 3 of the plant Arabidopsis
RT thaliana.";
RL Nature 408:820-822(2000).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=22954850; PubMed=14593172;
RA Yamada K., Lim J., Dale J.M., Chen H., Shinn P., Palm C.J.,
RA Southwick A.M., Wu H.C., Kim C.J., Nguyen M., Pham P.K., Cheuk R.F.,
RA Karlin-Newmann G., Liu S.X., Lam B., Sakano H., Wu T., Yu G.,
RA Miranda M., Quach H.L., Tripp M., Chang C.H., Lee J.M., Toriumi M.J.,
RA Chan M.M., Tang C.C., Onodera C.S., Deng J.M., Akiyama K., Ansari Y.,
RA Arakawa T., Banh J., Banno F., Bowser D., Brooks S.Y., Carninci P.,
RA Chao Q., Choy N., Enju A., Goldsmith A.D., Gurjal M., Hansen N.F.,
RA Hayaishizaki Y., Johnson-Hopson C., Hsuan V.W., Iida K., Karnes M.,
RA Khan S., Koesema E., Ishida J., Jiang P.X., Jones T., Kawai J.,
RA Kamiya A., Meyers C., Nakajima M., Narusaka M., Seki M., Sakurai T.,
RA Satou M., Tamse R., Vaysberg M., Wallender E.K., Wong C., Yamamura Y.,
RA Yuan S., Shinozaki K., Davis R.W., Theologis A., Ecker J.R.;
RT "Empirical analysis of transcriptional activity in the Arabidopsis
RT genome.";
RL Science 302:842-846(2003).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: Belongs to the L18E family of ribosomal proteins.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
DR EMBL; U15741; AAA69928.1; -.
DR EMBL; AC011620; AAF26138.1; -.
DR EMBL; AY037224; AAK59824.1; -.
DR EMBL; AY060533; RAL31164.1; -.
DR InterPro; IPR000039; Ribosomal_L18e.
DR Pfam; PF00828; Ribosomal_L18e; 1.
DR PROSITE; PS01106; RIBOSOMAL_L18E; FALSE_NEG.
KW Ribosomal protein.

```

FT CONFLICT 59 59 K -> R (IN REF. 1).  
SQ SEQUENCE 187 AA; 20926 MW; DDC72C02B12C8AB CRC64;

Query Match 100.0%; Score 26; DB 1; Length 187;  
Best Local Similarity 100.0%; Pred. No. 63;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EKAGG 5  
|||||  
Db 113 EKAGG 117

## RESULT 9

APT\_BIFLO STANDARD; PRT; 193 AA.  
AC Q8G6B5;  
DT 15-MAR-2004 (Rel. 43, Created)  
DT 15-MAR-2004 (Rel. 43, Last sequence update)  
DT 15-MAR-2004 (Rel. 43, Last annotation update)  
DE Adenine phosphoribosyltransferase (EC 2.4.2.7) (APRT).  
GN APT OR BL0731.  
OS Bifidobacterium longum.  
OC Bacteria; Actinobacteria; Actinobacteridae; Bifidobacteriales;  
OC Bifidobacteriaceae; Bifidobacterium.  
OX NCBI\_TaxID=216816;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=NCC 2705;  
RX MEDLINE=22294977; PubMed=12381787;  
RA Schell M.A., Karmirantzou M., Snel B., Vilanova D., Berger B.,  
RA Pessi G., Zwielen M.-C., Desiere F., Bork P., Delley M.,  
RA Pridmore R.D., Arigoni F.,  
RA "The genome sequence of Bifidobacterium longum reflects its adaptation  
RT to the human gastrointestinal tract";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:14422-14427(2002).  
CC -!- FUNCTION: Catalyzes a salvage reaction resulting in the formation  
CC of AMP, that is energetically less costly than de novo synthesis.  
CC -!- CATALYTIC ACTIVITY: AMP + diphosphate = adenine + 5-phospho-alpha-  
CC D-ribose 1-diphosphate.  
CC -!- PATHWAY: Purine salvage.  
CC -!- SUBUNIT: Homodimer (By similarity).  
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.  
CC -!- SIMILARITY: Belongs to the purine/pyrimidine  
CC phosphoribosyltransferase family.

-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
or send an email to [license@sb-sib.ch](mailto:license@sb-sib.ch)).  
-----

EMBL; AE014695; AAM24548.1; -;  
HAMAP; MF 00004; -; 1.  
InterPro: IPR005764; ADe\_phspho\_trans.  
InterPro: IPR002375; Pz\_Py\_rp\_transf.  
InterPro: IPR000836; PIRtransferase.  
Pfam; PF00156; Pribosyltran; 1.  
TIGRFAMs; TIGR01090; apt; 1.  
PROSITE; PS00103; PUR\_PYR\_PR\_TRANSFER; 1.  
Transferase; Glycosyltransferase; Purine salvage; Complete proteome.  
SEQUENCE 193 AA; 20251 MW; 43A0611C52F8EAF8 CRC64;

Query Match 100.0%; Score 26; DB 1; Length 193;  
Best Local Similarity 100.0%; Pred. No. 65;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EKAGG 5  
|||||  
Db 156 EKAGG 160

RESULT 10  
WRBA ECO57  
ID WRBA ECO57 STANDARD; PRT; 197 AA.  
AC Q8X4B4; Q8X4E8; Q9KXH7;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Flavoprotein wrba (Trp repressor binding protein).  
GN WRBA OR Z1423/Z1504 OR ECS1159.1/ECS1251.1.  
OS Escherichia coli O157:H7.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Escherichia.  
OX NCBI\_TaxID=83334;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=O157:H7 / RIMD 0509952;  
RX MEDLINE=20198780; PubMed=10734605;  
RA Makino K., Yokoyama K., Kubota Y., Yutsudo C.H., Kimura S.,  
RA Kurokawa K., Ishii K., Hattori M., Tatsuno I., Abe H., Iida T.,  
RA Yamamoto K., Ohnishi M., Hayaashi T., Yasunaga T., Honda T.,  
RA Sasaki C., Shinagawa H.;  
RT "Complete nucleotide sequence of the prophage VT2-Sakai carrying the  
RT verotoxin 2 genes of the enterohemorrhagic Escherichia coli O157:H7  
RT derived from the Sakai outbreak";  
RL Genes Genet. Syst. 74:227-239(1999).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;  
RX MEDLINE=21074935; PubMed=11206551;  
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,  
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,  
RA Postel G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,  
RA Grotbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,  
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,  
RA Welch R.A., Blattner F.R.;  
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7";  
RL Nature 409:529-533(2001).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=O157:H7 / RIMD 0509952;  
RX MEDLINE=21156231; PubMed=11258796;  
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,  
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Robe T.,  
RA Iida T., Takami H., Honda T., Sasaki C., Ogasawara N., Yasunaga T.,  
RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;  
RT "Complete genome sequence of enterohemorrhagic Escherichia coli  
RT O157:H7 and genomic comparison with a laboratory strain K-12";  
RL DNA Res. 8:11-22(2001).  
CC -!- FUNCTION: Seems to enhance the formation and/or stability of  
CC noncovalent complexes between the trp repressor protein and  
CC operator-bearing DNA (By similarity).  
CC -!- COFACTOR: Binds 1 FMN per monomer (By similarity).  
CC -!- SIMILARITY: Belongs to the wrba family.  
CC -!- SIMILARITY: Contains 1 flavodoxin-like domain.  
CC -!- CAUTION: Wrba is interrupted by the insertion of the prophage VT2-  
CC Sakai between positions 17 and 18.  
-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
or send an email to [license@sb-sib.ch](mailto:license@sb-sib.ch)).  
-----

EMBL; AP000422; BAA94098.1; -;  
EMBL; AE005294; AAG55552.1; ALT\_INIT.  
EMBL; AE005300; AAG55621.1; ALT\_TERM.  
EMBL; AP002554; -; NOT ANNOTATED\_CDS.  
HAMAP; MF 01017; -; 1.  
InterPro: IPR008254; Flav\_nitox\_synth.  
InterPro: IPR001226; Flavodoxin.  
Pfam; PF00258; Flavodoxin; 1.

```

DR PROSITE; PSS0902; FLAVODOXIN LIKE; 1.
KW Flavoprotein; FMN; Complete proteome.
FT INIT MET 0 0 BY SIMILARITY.
FT DOMAIN 3 188 FLAVODOXIN-LIKE.
SQ SEQUENCE 197 AA; 20703 MW; 7117EF93D3B72447 CRC64;

Query Match 100.0%; Score 26; DB 1; Length 197;
Best Local Similarity 100.0%; Pred. No. 66;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EKAGG 5
Db 48 EKAGG 52

RESULT 11
ID WRBA_ECOLI STANDARD; PRT; 197 AA.
AC P30849; P75890; P77543;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE Flavoprotein wrba (Trp repressor binding protein).
GN WRBA OR B1004 OR C1140 OR SF1008 OR S1077.
OS Escherichia coli,
OS Escherichia coli O6, and
OS Shigella flexneri.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562, 217992, 623;
RN [1] SEQUENCE FROM N.A. AND SEQUENCE OF 1-18.
RP SPECIES=E.coli; STRAIN=K12;
RC SPECIES=E.coli; STRAIN=K12;
RX MEDLINE=93296226; PubMed=8516330;
RA Yang W., Ni L., Somerville R.L.;
RT "A stationary-phase protein of Escherichia coli that affects the mode
RT of association between the trp repressor protein and operator-bearing
RT DNA".;
RL Proc. Natl. Acad. Sci. U.S.A. 90:5796-5800(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES=E.coli; STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-VIDES J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [3]
RP SEQUENCE FROM N.A.
RC SPECIES=E.coli; STRAIN=K12;
RX MEDLINE=97061202; PubMed=8905232;
RA Oshima T., Aiba H., Baba T., Fujita K., Hayaashi K., Honjo A.,
RA Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
RA Mori H., Motomura K., Nakamura Y., Nashimoto H., Nishio Y., Saito N.,
RA Sampei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
RA Yano M., Horiuchi T.;
RT "A 718-kb DNA sequence of the Escherichia coli K-12 genome
RT corresponding to the 12.7-28.0 min region on the linkage map.";
RL DNA Res. 3:137-155(1996).
RN [4]
RP SEQUENCE FROM N.A.
RC SPECIES=E.coli; STRAIN=O6:H1 / CFT073 / ATCC 700928;
RX MEDLINE=22388234; PubMed=12471157;
RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,
RA Rasko D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,
RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,
RA Mobley H.L.T., Donnenberg M.S., Blattner F.R.;
RT "Extensive mosaic structure revealed by the complete genome sequence
RT of uropathogenic Escherichia coli.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024(2002).

```

---

```

RN RP SEQUENCE OF 1-10, AND CHARACTERIZATION.
RC SPECIES=E.coli; STRAIN=JM101;
RX MEDLINE=98361966; PubMed=9694845;
RA Grandori R., Khalifah P., Boice J.A., Fairman R., Giovanielli K.,
RA Carey J.;
RT "Biochemical characterization of wrbA, founding member of a new family
RT of multimeric flavodoxin-like proteins.";
RL J. Biol. Chem. 273:20960-20966(1998).
RN [6]
RP SEQUENCE FROM N.A.
RC SPECIES=S.flexneri; STRAIN=301 / Serotype 2a;
RX MEDLINE=2272406; PubMed=12384590;
RA Jin Q., Yuan Z., Xu J., Wang Y., Shen Y., Lu W., Wang J., Liu H.,
RA Yang J., Yang F., Zhang X., Zhang J., Yang G., Wu H., Qu D., Dong J.,
RA Sun L., Xue Y., Zhao A., Gao Y., Zhu J., Kan B., Ding K., Chen S.,
RA Cheng H., Yao Z., He B., Chen R., Ma D., Qiang B., Wen Y., Hou Y.,
RA Yu J.;
RT "Genome sequence of Shigella flexneri 2a: insights into pathogenicity
RT through comparison with genomes of Escherichia coli K12 and O157.";
RL Nucleic Acids Res. 30:4432-4441(2002).
RN [7]
RP SEQUENCE FROM N.A.
RC SPECIES=S.flexneri; STRAIN=2457T / ATCC 700930 / Serotype 2a;
RX MEDLINE=22590274; PubMed=12704152;
RA Wei J., Goldberg M.B., Burland V., Venkatesan M.M., Deng W.,
RA Fournier G., Mayhew G.F., Plunkett G. III, Rose D.J., Darling A.,
RA Mau B., Perna N.T., Payne S.M., Runyen-Janecky L.J., Zhou S.,
RA Schwartz D.C., Blattner F.R.;
RT "Complete genome sequence and comparative genomics of Shigella
RT flexneri serotype 2a strain 2457T.";
RL Infect. Immun. 71:2775-2786(2003).
CC -1- FUNCTION: Seems to enhance the formation and/or stability of
CC noncovalent complexes between the trp repressor protein and
CC operator-bearing DNA. However, wrbA does not specifically
CC influence the affinity or mode of binding of trpR to its operator.
CC It seems possible that the association may have a structural,
CC rather than functional, basis. wrbA alone does not interact with
CC the operator-bearing DNA; the trpR/trpO complex is probably the
CC major target species. The wrbA protein may function as an
CC accessory element in blocking trpR-specific transcriptional
CC processes that might be physiologically disadvantageous in the
CC stationary phase of the bacterial life cycle.
CC -1- COFACTOR: Binds 1 FMN per monomer.
CC -1- SUBUNIT: Homodimer and homotetramer; in equilibrium.
CC -1- SIMILARITY: Belongs to the wrbA family. domain.
CC -1- SIMILARITY: Contains 1 Flavodoxin-like domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M99166; AAA24759.1; -
CC EMBL; A8000202; AAC74089.1; -
CC EMBL; D90738; BAA35781.1; -
CC EMBL; D90737; BAA35771.1; -
CC EMBL; AE016758; AAN79608.1; -
CC EMBL; AE015129; AAN42634.1; -
CC EMBL; AE016981; AAP16518.1; -
CC FIR; B64842; B64842.
CC SWISS-2DPAGE; P30849; COLI.
CC EcoGene; EGI1540; wrbA.
CC HAMAP; MF 01017; -; 1.
CC InterPro; IPR008254; Flav nitox synth.
CC InterPro; IPR001226; Flavodoxin.
CC Pfam; PF00258; flavodoxin; 1.
CC PROSITE; PSS0902; FLAVODOXIN LIKE; 1.
KW Flavoprotein; FMN; Complete proteome.
FT INIT MET 0 0

```

```

FT DOMAIN 3 188 FLAVODOXIN-LIKE
FT CONFLICT 141 141 G -> A (IN REF. 1).
SQ SEQUENCE 197 AA; 20714 MW; 7B48C7CCC2B71747 CRC64;

Query Match
Best Local Similarity 100.0%; Score 26; DB 1; Length 197;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 EKAGG 5
Db 48 EKAGG 52

RESULT 12
HB2G HUMAN
ID HB2G HUMAN STANDARD; PRT; 198 AA.
AC P01911;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE HLA class II histocompatibility antigen, DR2.2/DR2.2 beta chain
DE (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE.
RX MEDLINE=82096757; PubMed=6947956;
RA Kratzin H., Yang C.-Y., Gotz H., Pauly E., Kolbel S., Egert G.,
RA Thines F.P., Wernet P., Altevogt P., Hilschmann N.;
RA "Primary structure of class II human histocompatibility antigens. 1st
RA communication. Amino acid sequence of the N-terminal 198 residues of
RA the beta chain of a HLA-Dw2,DR2,2-allele antigen."
RA Hoppe-Sejler's Z. Physiol. Chem. 362:1665-1669(1981).
CC -!- MISCELLANEOUS: THE CHAIN SHOWN CONSTITUTED ABOUT 70% OF A POOL OF
CC AT LEAST SEVEN SIMILAR BETA CHAINS.
DR HSSP; P13758; IDLH.
DR GO; GO:0016020; C:membrane; NAS.
DR GO; GO:0045012; F:MHC class II receptor activity; NAS.
DR GO; GO:0006955; P:immune response; NAS.
DR InterPro; IPR007110; IG-like.
DR InterPro; IPR003597; IG cl.
DR InterPro; IPR003006; IG MHC.
DR InterPro; IPR000353; MHC_II_beta.
DR Pfam; PF00047; Ig; 1.
DR Pfam; PF00969; MHC_II_beta; 1.
DR PRODOM; PD000328; MHC_II_beta; 1.
DR SMART; SM00407; IGcl; 1.
DR PROSITE; PS00835; IG LIKE; 1.
DR PROSITE; PS00290; IG MHC; 1.
KW MHC II; Transmembrane; Glycoprotein.
FT DOMAIN 1 94
FT DOMAIN 95 187 EXTRACELLULAR BETA-1.
FT DOMAIN 188 197 EXTRACELLULAR BETA-2.
FT TRANSMEM 198 >198 CONNECTING PEPTIDE.
FT DISULFID 15 79
FT DISULFID 117 172 BY SIMILARITY.
FT CARBOHYD 19 19 BY SIMILARITY.
FT NON TER 198 198 N-LINKED (GLCNAC. . .).
SQ SEQUENCE 198 AA; 22965 MW; 18199D91A4D737EC CRC64;

Query Match
Best Local Similarity 100.0%; Score 26; DB 1; Length 198;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 EKAGG 5
Db 138 EKAGG 142

RESULT 13
P25_SCHPO
ID P25_SCHPO STANDARD; PRT; 202 AA.
AC P30821;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE P25 protein (Brefeldin A resistance protein).
GN OBR1 OR SPAC3C7.14C.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 36-56.
RX MEDLINE=93078747; PubMed=1448080;
RA Toda T., Shimanuki M., Saka Y., Yamano H., Adachi Y.,
RA Shirakawa M., Kyogoku Y., Yanagida M.;
RA "Fission yeast papi-dependent transcription is negatively regulated
RA by an essential nuclear protein, crml."
RA Mol. Cell. Biol. 12:5474-5484(1992).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=95014161; PubMed=7929079;
RA Turi T.G., Webster P., Rose J.K.;
RA "Brefeldin A sensitivity and resistance in Schizosaccharomyces pombe.
RA Isolation of multiple genes conferring resistance."
RA J. Biol. Chem. 269:24229-24236(1994).
RN [3]
RP SEQUENCE FROM N.A.
RX STRAIN=972; PubMed=11859360;
RX MEDLINE=21848401; PubMed=11859360;
RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Volckaert G., Aert R., Robben J., Grymoprez B.,
RA Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Lebrach H., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Cerutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
RA "The genome sequence of Schizosaccharomyces pombe."
RA Nature 415:871-880(2002).
CC -!- FUNCTION: UNKNOWN. TARGET OF PAPI TRANSCRIPTION FACTOR. CONFERS
CC -!- SUBUNIT: Homodimer.
CC -!- SIMILARITY: Belongs to the wrba family.
CC -!- SIMILARITY: Contains 1 flavodoxin-like domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D13038; BAA02370.1; -.
DR EMBL; X73558; CAA51956.1; -.

```

```

DR EMBL; Z99568; CAB16744.1; -.
DR PIR; A45029; A45029.
DR GenDB SPombe; SPAC3C7.14c; -.
DR InterPro; IPR008254; Flav_nitox_synth.
DR Pfam; PF00258; flavodoxin; 1.
DR PROSITE; PS50902; FLAVODOXIN_LIKE; 1.
FT DOMAIN 7 135 FLAVODOXIN-LIKE.
SQ SEQUENCE 202 AA; 21899 MW; 54E1DB3A8F7E9DB2 CRC64;

Query Match 100.0%; Score 26; DB 1; Length 202;
Best Local Similarity 100.0%; Pred. No. 68;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EKAGG 5
Db 30 EKAGG 34

RESULT 14
CMEI_BACSU STANDARD; PRT; 205 AA.
AC P39694;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE COME operon protein 1.
GN COMEA OR COME1 OR BSU25590.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]

SEQUENCE FROM N.A.
RP MEDLINE=95058187; PubMed=79698523;
RA Hahn J., Inamine G., Kozlov Y., Dubnau D.A.;
RT "Characterization of comE, a late competence operon of Bacillus
RL subtilis required for the binding and uptake of transforming DNA.";
RN [2]

SEQUENCE FROM N.A.
RP STRAIN=168 / JH642;
RX MEDLINE=97124195; PubMed=8969508;
RA Mizuno M., Masuda S., Takemaru K.-I., Hosono S., Sato T., Takeuchi M.,
RA Kobayashi Y.;
RT "Systematic sequencing of the 283 kb 210 degrees-232 degrees region of
RT the Bacillus subtilis genome containing the skin element and many
RL sporulation genes.";
RN [3]

Microbiology 142:3103-3111(1996).
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA Boriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Connerston I.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Ghim S.Y., Glaser P., Goffeau A., Gollightly E.J., Grandi G.,
RA Guisepi G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA Kobayashi Y., Koetter P., Koningsstein G., Krohn S., Kumano M.,
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogiwara K., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA Presecan E., Pujic P., Furnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadate Y.,
RA Sato T., Scalan E., Schleisch S., Schroeter R., Scoffone F.,
RA Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,

```

```

RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzenecker T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;
RT "The complete genome sequence of the Gram-positive bacterium Bacillus
RT subtilis";
RL Nature 390:249-256(1997).

RN [4]
CHARACTERIZATION.
RP MEDLINE=95286482; PubMed=7768800;
RX Inamine G.S., Dubnau D.;
RA "ComeA, a Bacillus subtilis integral membrane protein required for
RT genetic transformation, is needed for both DNA binding and
RT transport.";
RL J. Bacteriol. 177:3045-3051(1995).
CC -!- FUNCTION: Needed for both DNA binding and transport. It is
CC absolutely required for the uptake of transforming DNA but not for
CC binding. Its role in binding may be indirect.
CC -!- SUBCELLULAR LOCATION: Type II membrane protein. Inner membrane.
CC -!- SIMILARITY: TO E.COLI YEAV AND H.INFLUENZAE HI1008.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).

EMBL; LI5202; AAC36905.1; -
EMBL; D84432; BAA12452.1; -
EMBL; Z99117; CAB14501.1; -
PIR; S39863; S39863.
Subtilist; BG10480; COMEA.
InterPro; IPR004787; COMEA.
InterPro; IPR004509; COMEA_HHH.
InterPro; IPR000445; Hhh.
InterPro; IPR003583; HHH_1.
SMART; PF00633; HHH; 2.
SMART; SM00278; HHH1; 2.
TIGRFAMs; TIGR01259; comE; 1.
TIGRFAMs; TIGR00426; TIGR00426; 1.
Competence; Transmembrane; Complete proteome.
DOMAIN 1 9 CYTOPLASMIC (POTENTIAL).
TRANSMEM 10 28 POTENTIAL.
FT DOMAIN 29 205 EXTRACELLULAR (POTENTIAL).
SQ SEQUENCE 205 AA; 21769 MW; 4E3B17FAF982F9DE CRC64;

Query Match 100.0%; Score 26; DB 1; Length 205;
Best Local Similarity 100.0%; Pred. No. 69;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EKAGG 5
Db 89 EKAGG 93

RESULT 15
RNC_LISIN
ID RNC_LISIN STANDARD; PRT; 229 AA.
AC Q92AK3;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Ribonuclease III (EC 3.1.26.3) (RNase III).
GN RNC OR RNCs OR LIN1919.
OS Listeria innocua.
OC Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
OX NCBI_TaxID=1642;
RN [1]
RP SEQUENCE FROM N.A.

```

```

RC STRAIN-CLIP 11262 / Serovar 6a;
RX MEDLINE=21537279; PubMed=11679669;
RA Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,
RA Baquero F., Berche P., Bloeker H., Brandt P., Chakraborty T.,
RA Charbit A., Chetouani F., Couve E., de Daruvar A., Dehoux P.,
RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussauget O.,
RA Entian K.-D., Frihi H., Garcia-del Portillo F., Garrido P.,
RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
RA Jones L.-M., Kaerst U., Kretz J., Kuhn M., Kunst F., Kurupkat G.,
RA Madueno E., Mailcounam A., Mata Vicente J., Ng E., Nedjari H.,
RA Nordsiek G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,
RA Remmel B., Rose M., Schluter T., Simoes N., Tierrez A.,
RA Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.;
RT "Comparative genomics of Listeria species.";
RL Science 294:849-852(2001).
CC -1- FUNCTION: Digests double-stranded RNA. Involved in the processing
CC of ribosomal RNA precursors and of some mRNAs (By similarity).
CC -1- CATALYTIC ACTIVITY: Endonucleolytic cleavage to 5'-
CC phosphomonoester.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- SIMILARITY: Contains 1 DBRM (double-stranded RNA-binding) domain.
CC -1- SIMILARITY: Contains 1 RNase III domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AL596170; CAC97149.1; -.
DR PIR; AE1672; AE1672.
DR List1List; LIN01919; -.
DR HAMAP; MF 00104; -.
DR InterPro; IPR001159; DS_RBD.
DR InterPro; IPR000999; RNase_III.
DR Pfam; PF00035; dsrm; 1.
DR Pfam; PF00636; Ribonuclease_3; 1.
DR SMART; SM00358; DSRM; 1.
DR SMART; SM00535; RIBOC; 1.
DR PROSITE; PS01337; DS_RBD; 1.
DR PROSITE; PS00517; RNase_3.1; 1.
DR PROSITE; PS0142; RNase_3.2; 1.
DR Hydrolase; Nuclease; Endonuclease; RNA-binding; Complete proteome.
FT DOMAIN 4 133 RNase III.
FT DOMAIN 159 228
FT SEQUENCE 229 AA; 25932 MW; 941D48C27A41DFC4 CRC64;
SQ
Query Match 100.0%; Score 26; DB 1; Length 229;
Best Local Similarity 100.0%; Pred. No. 77;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 EKAGG 5
Db 106 EKAGG 110

```





GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 18, 2004, 04:00:08 ; Search time 17.7451 Seconds  
(without alignments)  
88.903 Million cell updates/sec

Title: US-10-615-959-44

Perfect score: 26

Sequence: 1 EKAGG 5

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL 25:\*

- 1: sp\_archaea:\*
- 2: sp\_bacteria:\*
- 3: sp\_fungi:\*
- 4: sp\_human:\*
- 5: sp\_invertebrate:\*
- 6: sp\_mammal:\*
- 7: sp\_mhc:\*
- 8: sp\_organelle:\*
- 9: sp\_phage:\*
- 10: sp\_plant:\*
- 11: sp\_rodent:\*
- 12: sp\_virus:\*
- 13: sp\_vertebrate:\*
- 14: sp\_unclassified:\*
- 15: sp\_rvirus:\*
- 16: sp\_bacteriap:\*
- 17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	26	100.0	56	11	Q9QXC2
2	26	100.0	63	11	Q9QXA8
3	26	100.0	68	10	Q9M892
4	26	100.0	74	16	Q822T5
5	26	100.0	84	2	Q9ZHV7
6	26	100.0	84	2	Q9APD1
7	26	100.0	90	16	Q7U5V8
8	26	100.0	95	17	Q8XP7
9	26	100.0	118	10	Q851J5
10	26	100.0	120	16	Q8XGX3
11	26	100.0	126	16	Q8YXP8
12	26	100.0	126	16	Q9L192
13	26	100.0	128	12	O11718
14	26	100.0	128	12	O11702
15	26	100.0	128	12	O11779
16	26	100.0	128	12	O11700

17	26	100.0	128	12	O11701
18	26	100.0	129	10	Q945D7
19	26	100.0	130	2	O85956
20	26	100.0	134	11	Q8BXK1
21	26	100.0	145	17	Q97VH5
22	26	100.0	146	16	Q8U9L0
23	26	100.0	150	10	Q7X6D7
24	26	100.0	150	16	Q8R8A6
25	26	100.0	151	16	Q92M94
26	26	100.0	153	16	Q92GY5
27	26	100.0	155	11	Q9DA76
28	26	100.0	156	16	Q8YHM1
29	26	100.0	156	16	Q92QF1
30	26	100.0	156	16	Q8G090
31	26	100.0	156	17	Q8ZT50
32	26	100.0	157	16	Q98N38
33	26	100.0	159	16	Q92Z42
34	26	100.0	161	16	Q89JA3
35	26	100.0	162	10	Q9LKN3
36	26	100.0	164	10	Q84U89
37	26	100.0	164	16	Q9JS31
38	26	100.0	165	16	Q9JZM6
39	26	100.0	170	5	Q8I998
40	26	100.0	172	16	Q7V7D8
41	26	100.0	176	16	Q7WLR6
42	26	100.0	177	16	Q8RHG3
43	26	100.0	179	5	Q27706
44	26	100.0	182	16	Q7V1T3
45	26	100.0	184	10	Q852G1

#### ALIGNMENTS

RESULT 1

Q9QXC2 PRELIMINARY; PRT; 56 AA.

AC Q9QXC2; 01-MAY-2000 (TREMBLrel. 13, Created)

DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)

DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)

DE Fructose-1,6-bisphosphatase (EC 3.1.3.11) (fragment).

GN FBPI OR FBPASE 1 GENE.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus; Mus.

OX NCBI\_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Liver;

RX MEDLINE=21147931; PubMed=11250076;

RA Stein S., Liehr T., Eschrich K.;

RT "Characterization of the mouse liver fructose-1,6-bisphosphatase gene.";

RL Gene 264.215-224(2001).

DR EMBL; AJ242922; CAB65257.1; --

DR HSSP; P00636; IFRP.

DR MGD; MGI:95492; Fbp1.

DR GO; GO:0042132; F:fructose-bisphosphatase activity; IEA.

DR GO; GO:0016787; F:hydrolase activity; IEA.

DR GO; GO:0042578; F:phosphoric ester hydrolase activity; IEA.

DR GO; GO:0005975; P:carbohydrate metabolism; IEA.

DR InterPro; IPR000146; In\_FB\_phptase.

DR Pfam; PF00316; FBPAse; 1.

DR ProDom; PD001491; In\_FB\_phptase; 1.

KW Hydrolase.

KW NON\_TER

FT NON\_TER 1 56

FT NON\_TER 56

SQ SEQUENCE 56 AA; 6162 MW; B25814CECE14C462 CRC64;

Query Match 100.0%; Score 26; DB 11; Length 56;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

[illegible]

```

Db          30 EKAGG 34

RESULT 5
Q9ZHY7      PRELIMINARY;          PRT;      84 AA.
AC          Q9ZHY7;
AC          Q9ZHY7;
DT          01-MAY-1999 (TrEMBLrel. 10, Created)
DT          01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT          01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE          Carboxysome polypeptide.
OS          Thiobacillus intermedius.
OC          Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC          Comamonadaceae; Thiomonas.
OX          NCBI_TaxID=926;
RN          [1]
RP          SEQUENCE FROM N.A.
RC          STRAIN=K12;
RA          Shively J.M.;
RL          Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
DR          EMBL; AF046933; AAD02449.1; -.
DR          InterPro; IPR004992; EutN_CcmL.
DR          Pfam; PF03319; EutN_CcmL; 1.
DR          ProDom; PD012510; EutN_CcmL; 1.
SQ          SEQUENCE 84 AA; 8879 MW; 4CF8350D591A5C7D CRC64;

Query Match          100.0%; Score 26; DB 2; Length 84;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy          1 EKAGG 5
Db          29 EKAGG 33

RESULT 6
Q9APD1
ID          Q9APD1      PRELIMINARY;          PRT;      84 AA.
AC          Q9APD1;
DT          01-JUN-2001 (TrEMBLrel. 17, Created)
DT          01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT          01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE          Carboxysome peptide ORF A.
OS          Thiobacillus denitrificans.
OC          Bacteria; Proteobacteria; Betaproteobacteria; Thiobacillus.
OX          NCBI_TaxID=36861;
RN          [1]
RP          SEQUENCE FROM N.A.
RC          STRAIN=ATCC25259;
RA          Cannon G.C., Heinrich S., Soyer F., Shively J.M.;
RL          "The putative carboxysome operon of Thiobacillus denitrificans.";
RL          Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
DR          EMBL; AF307030; AAG60039.1; -.
DR          InterPro; IPR004992; EutN_CcmL.
DR          Pfam; PF03319; EutN_CcmL; 1.
DR          ProDom; PD012510; EutN_CcmL; 1.
SQ          SEQUENCE 84 AA; 9110 MW; B975A85088BA838C CRC64;

Query Match          100.0%; Score 26; DB 2; Length 84;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy          1 EKAGG 5
Db          29 EKAGG 33

RESULT 7
Q7U5V8
ID          Q7U5V8      PRELIMINARY;          PRT;      90 AA.
AC          Q7U5V8;
DT          01-OCT-2003 (TrEMBLrel. 25, Created)
DT          01-OCT-2003 (TrEMBLrel. 25, Last sequence update)

```

---

```

DT          01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE          Conserved hypothetical.
GN          SYN1583.
OS          Synecococcus sp. (strain WH8102).
OC          Bacteria; Cyanobacteria; Chroococcales; Synecococcus.
OX          NCBI_TaxID=84588;
RN          [1]
RP          SEQUENCE FROM N.A.
RC          MEDLINE=22825697; PubMed=12917641;
RA          Palenik B., Brahamsha B., Larimer F.W., Land M., Hauser L., Chain P.,
RA          Lamerik J., Regala W., Allen E.E., McCarran J., Paulsen I.,
RA          Dufresne A., Partensky F., Webb E.A., Waterbury J.;
RT          "The genome of a motile marine Synecococcus.";
RL          Nature 424:1037-1042(2003).
DR          EMBL; BX569693; CAE08098.1; -.
KW          Hypothetical protein; Complete proteome.
SQ          SEQUENCE 90 AA; 10321 MW; E60FC6B1902B3D63 CRC64;

Query Match          100.0%; Score 26; DB 16; Length 90;
Best Local Similarity 100.0%; Pred. No. 2.4e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy          1 EKAGG 5
Db          13 EKAGG 17

RESULT 8
Q8PXP7      PRELIMINARY;          PRT;      95 AA.
AC          Q8PXP7;
DT          01-OCT-2002 (TrEMBLrel. 22, Created)
DT          01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT          01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE          Glucose-1-phosphate thymidyltransferase (EC 2.7.7.24).
GN          MM1169.
OS          Methanosarcina mazei (Methanosarcina frisia).
OC          Archaea; Euryarchaeota; Euryarchaeota orders incertae sedis;
OC          Methanosarcinales; Methanosarcinaceae; Methanosarcina.
OX          NCBI_TaxID=2209;
RN          [1]
RP          SEQUENCE FROM N.A.
RC          STRAIN=Goe1 / Goe1 / ATCC BAA-199 / DSM 3647 / OCM 88;
RX          MEDLINE=22120827; PubMed=12125824;
RA          Deppenmeier U., Johann A., Hartsch T., Merkl R., Schmitz R.A.,
RA          Martinez-Arias R., Henne A., Wierzer A., Baeumer S., Jacobi C.,
RA          Brueggemann H., Lienard T., Christmann A., Boemecke M., Steckel S.,
RA          Bhattacharyya A., Lykidis A., Overbeek R., Klenk H.-P., Gunsalus R.P.,
RA          Fritz H.-J., Gottschalk G.;
RT          "The genome of Methanosarcina mazei: evidence for lateral gene
RT          transfer between Bacteria and Archaea.";
RL          J. Mol. Microbiol. Biotechnol. 4:453-461(2002).
DR          EMBL; AE013347; AAM30865.1; -.
DR          GO; GO:0008879; F:glucose-1-phosphate thymidyltransferase a. . .; IEA.
DR          GO; GO:0016740; F:transferase activity; IEA.
DR          GO; GO:0009058; P:biosynthesis; IEA.
DR          InterPro; IPR005835; NTP transferase; 1.
DR          Pfam; PF00483; NTP transferase; 1.
KW          Transferase; Nucleotidyltransferase; Complete proteome.
SQ          SEQUENCE 95 AA; 10409 MW; 10D22BFF2445144E CRC64;

Query Match          100.0%; Score 26; DB 17; Length 95;
Best Local Similarity 100.0%; Pred. No. 2.5e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy          1 EKAGG 5
Db          81 EKAGG 85

RESULT 9
Q851J5      PRELIMINARY;          PRT;      118 AA.
ID          Q851J5

```

```

AC Q851J5;
DT 01-JUN-2003 (T-EMBLrel. 24, Created)
DT 01-JUN-2003 (T-EMBLrel. 24, Last sequence update)
DT 01-JUN-2003 (T-EMBLrel. 24, Last annotation update)
DE Hypothetical protein.
GN OSJNB0021011.17.
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Erihartoideae; Oryzaceae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Nipponbare;
RA Buell C.R., Yuan Q., Ouyang S., Liu J., Gansberger K., Jones K.M.,
RA Overton II L.L., Tsitrin T., Kim M.M., Bera J.J., Jin S.S.,
RA Fadrosch D.W., Tallon L.J., Koo H., Zismann V., Hsiao J., Blunt S.,
RA Vanaken S.S., Riedmuller S.B., Utterback T.T., Feldblyum T.V.,
RA Yang Q.Q., Haas B.J., Suh B.B., Peterson J.J., Quackenbush J.,
RA White O., Salzberg S.L., Fraser C.M.; D89C1124A459C63 CRC64;
RT "Oryza sativa chromosome 3 BAC OSJNB0021011 genomic sequence.";
RL Submitted (FEB-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC120508; AAC38495.1; -.
KW Hypothetical protein.
SQ SEQUENCE 118 AA; 12495 MW; D89C1124A459C63 CRC64;

Query Match 100.0%; Score 26; DB 10; Length 118;
Best Local Similarity 100.0%; Pred. No. 3.2e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EKAGG 5
Db 95 EKAGG 99

RESULT 10
Q8XGX3 PRELIMINARY; PRT; 120 AA.
AC Q8XGX3;
DT 01-MAR-2002 (T-EMBLrel. 20, Created)
DT 01-MAR-2002 (T-EMBLrel. 20, Last sequence update)
DT 01-JUN-2003 (T-EMBLrel. 24, Last annotation update)
DE Hypothetical protein (Putative cytoplasmic protein).
GN T4475 OR STM441 OR SY4780.
OS Salmonella typhi, and
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=601, 602;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=S.typhi; STRAIN=Ty2 / ATCC 700931;
RX MEDLINE=22531367; PubMed=12644504;
RA Deng W., Liou S.-R., Plunkett G. III, Mayhew G.F., Rose D.J.,
RA Burland V., Kodoyianni V., Schwartz D.C., Blattner F.R.;
RT "Comparative genomics of Salmonella enterica serovar Typhi strains Ty2
and CT18.";
RL J. Bacteriol. 185:2330-2337(2003).
RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES=S.typhimurium; STRAIN=LT2 / SGSC1412 / ATCC 700720;
RX MEDLINE=21534948; PubMed=11677609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.;
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
LT2.";
RL Nature 413:852-856(2001).
RN [3]
RP SEQUENCE FROM N.A.
RC SPECIES=S.typhi; STRAIN=CT18;

```

```

RX MEDLINE=21534947; PubMed=11677608;
RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebaihia M.,
RA Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,
RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
RA Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,
RA Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,
RA Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,
RA Whitehead S., Barrall B.G.;
RT "Complete genome sequence of a multiple drug resistant Salmonella
enterica serovar Typhi CT18.";
RL Nature 413:848-852(2001).
DR EMBL; AB016849; AA071922.1; -.
DR EMBL; AF008908; AAL23261.1; -.
DR EMBL; AL627283; CAD06901.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 120 AA; 12469 MW; 11DF228CDF390A24 CRC64;

Query Match 100.0%; Score 26; DB 16; Length 120;
Best Local Similarity 100.0%; Pred. No. 3.3e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EKAGG 5
Db 23 EKAGG 27

RESULT 11
Q8YXP8 PRELIMINARY; PRT; 126 AA.
AC Q8YXP8;
DT 01-MAR-2002 (T-EMBLrel. 20, Created)
DT 01-MAR-2002 (T-EMBLrel. 20, Last sequence update)
DT 01-JUN-2003 (T-EMBLrel. 24, Last annotation update)
DE Hypothetical protein All1164.
GN ALL1164.
OS Anabaena sp. (strain PCC 7120).
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
OX NCBI_TaxID=103690;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21595285; PubMed=11759840;
RA Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,
RA Watanabe A., Iriguchi M., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
RA Nakazaki N., Shimpo S., Sugimoto M., Takazawa M., Yamada M.,
RA Yasuda M., Tabata S.;
RT "Complete genomic sequence of the filamentous nitrogen-fixing
cyanobacterium Anabaena sp. strain PCC 7120.";
RL DNA Res. 8:205-213(2001).
DR EMBL; AF003584; BAB73121.1; -.
DR PIR; A11951; A11951.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 126 AA; 14192 MW; 51EDF6752FB60723 CRC64;

Query Match 100.0%; Score 26; DB 16; Length 126;
Best Local Similarity 100.0%; Pred. No. 3.4e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EKAGG 5
Db 24 EKAGG 28

RESULT 12
Q9L192 PRELIMINARY; PRT; 126 AA.
AC Q9L192;
DT 01-OCT-2000 (T-EMBLrel. 15, Created)
DT 01-OCT-2000 (T-EMBLrel. 15, Last sequence update)
DT 01-MAR-2003 (T-EMBLrel. 23, Last annotation update)
DE Hypothetical protein SCO7383.
GN SCO7383 OR SC1068.10.

```

OS Streptomyces coelicolor.  
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 OC Streptomycineae; Streptomycetaceae; Streptomyces.  
 NCBI\_TaxID=1902;  
 RN [1]\_SEQUENCE FROM N.A.  
 RP STRAIN=A3(2) / M145;  
 RC MEDLINE=21996410; PubMed=12000953;  
 RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,  
 RA Thomson N.R., James K.D., Harris D.E., Ouail M.A., Kieser H.,  
 RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,  
 RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,  
 RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neill S.,  
 RA Rabinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,  
 RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,  
 RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,  
 RA Hopwood D.A.;  
 RT Complete genome sequence of the model actinomycete Streptomyces  
 coelicolor A3(2).";  
 RL Nature 417:141-147(2002).  
 DR EMBL; AL939131; CAB76284.1; --  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 126 AA; 13709 MW; 8B36C736317306BD CRC64;

Query Match 100.0%; Score 26; DB 16; Length 126;  
 Best Local Similarity 100.0%; Pred. No. 3.4e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EKAGG 5  
 |||||  
 Db 71 EKAGG 75

## RESULT 13

O11718  
 ID O11718 PRELIMINARY; PRT; 128 AA.  
 AC O11718;  
 DT 01-JUL-1997 (TrEMBLrel. 04, Created)  
 DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Phosphoprotein (Fragment).  
 OS Measles virus.  
 OC Viruses; ssRNA negative-strand viruses; Mononegavirales;  
 OC Paramyxoviridae; Paramyxovirinae; Morbilliviruses.  
 NCBI\_TaxID=11234;  
 RN [1]\_SEQUENCE FROM N.A.  
 RP STRAIN=SSPE93;  
 RC MEDLINE=97278133; PubMed=9131467;  
 RA Yamaguchi S.;  
 RT Identification of 3 lineages of wild measles virus by the nucleotide  
 sequence analysis of N, P, M, F and L gene for 10 years in Japan.";  
 RL J. Med. Virol. 52:113-120(1997).  
 DR EMBL; D87540; BAA20201.1; --  
 DR GO; GO:0003723; F:RNA binding; IEA.  
 DR GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.  
 DR GO; GO:0006350; P:transcription; IEA.  
 DR InterPro; IPR002581; Morbilli\_P.  
 DR Pfam; PF01647; Morbilli\_P; 1.  
 FT NON\_TER 128 128  
 SQ SEQUENCE 128 AA; 13335 MW; 62CD956627A9389B CRC64;

Query Match 100.0%; Score 26; DB 12; Length 128;  
 Best Local Similarity 100.0%; Pred. No. 3.5e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EKAGG 5  
 |||||  
 Db 53 EKAGG 57

## RESULT 14

O11702  
 ID O11702 PRELIMINARY; PRT; 128 AA.  
 AC O11702;  
 DT 01-JUL-1997 (TrEMBLrel. 04, Created)  
 DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Phosphoprotein (Fragment).  
 OS Measles virus.  
 OC Viruses; ssRNA negative-strand viruses; Mononegavirales;  
 OC Paramyxoviridae; Paramyxovirinae; Morbilliviruses.  
 NCBI\_TaxID=11234;  
 RN [1]\_SEQUENCE FROM N.A.  
 RP STRAIN=84-Y; PubMed=9131467;  
 RC MEDLINE=97278133; PubMed=9131467;  
 RA Yamaguchi S.;  
 RT Identification of 3 lineages of wild measles virus by the nucleotide  
 sequence analysis of N, P, M, F and L gene for 10 years in Japan.";  
 RL J. Med. Virol. 52:113-120(1997).  
 DR EMBL; D87540; BAA20201.1; --  
 DR GO; GO:0003723; F:RNA binding; IEA.  
 DR GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.  
 DR GO; GO:0006350; P:transcription; IEA.  
 DR InterPro; IPR002581; Morbilli\_P.  
 DR Pfam; PF01647; Morbilli\_P; 1.  
 FT NON\_TER 128 128  
 SQ SEQUENCE 128 AA; 13308 MW; CF2A0AFC3673F78 CRC64;

Query Match 100.0%; Score 26; DB 12; Length 128;  
 Best Local Similarity 100.0%; Pred. No. 3.5e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EKAGG 5  
 |||||  
 Db 53 EKAGG 57

ID O11702 PRELIMINARY; PRT; 128 AA.  
 AC O11702;  
 DT 01-JUL-1997 (TrEMBLrel. 04, Created)  
 DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Phosphoprotein (Fragment).  
 OS Measles virus.  
 OC Viruses; ssRNA negative-strand viruses; Mononegavirales;  
 OC Paramyxoviridae; Paramyxovirinae; Morbilliviruses.  
 NCBI\_TaxID=11234;  
 RN [1]\_SEQUENCE FROM N.A.  
 RP STRAIN=84-Y; PubMed=9131467;  
 RC MEDLINE=97278133; PubMed=9131467;  
 RA Yamaguchi S.;  
 RT Identification of 3 lineages of wild measles virus by the nucleotide  
 sequence analysis of N, P, M, F and L gene for 10 years in Japan.";  
 RL J. Med. Virol. 52:113-120(1997).  
 DR EMBL; D87524; BAA20185.1; --  
 DR GO; GO:0003723; F:RNA binding; IEA.  
 DR GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.  
 DR GO; GO:0006350; P:transcription; IEA.  
 DR InterPro; IPR002581; Morbilli\_P.  
 DR Pfam; PF01647; Morbilli\_P; 1.  
 FT NON\_TER 128 128  
 SQ SEQUENCE 128 AA; 13298 MW; E205DAAFC3673F7B CRC64;

Query Match 100.0%; Score 26; DB 12; Length 128;  
 Best Local Similarity 100.0%; Pred. No. 3.5e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EKAGG 5  
 |||||  
 Db 53 EKAGG 57

## RESULT 15

O11779  
 ID O11779 PRELIMINARY; PRT; 128 AA.  
 AC O11779;  
 DT 01-JUL-1997 (TrEMBLrel. 04, Created)  
 DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Phosphoprotein (Fragment).  
 OS Measles virus.  
 OC Viruses; ssRNA negative-strand viruses; Mononegavirales;  
 OC Paramyxoviridae; Paramyxovirinae; Morbilliviruses.  
 NCBI\_TaxID=11234;  
 RN [1]\_SEQUENCE FROM N.A.  
 RP STRAIN=84-K;  
 RC MEDLINE=97278133; PubMed=9131467;  
 RA Yamaguchi S.;  
 RT Identification of 3 lineages of wild measles virus by the nucleotide  
 sequence analysis of N, P, M, F and L gene for 10 years in Japan.";  
 RL J. Med. Virol. 52:113-120(1997).  
 DR EMBL; D87751; BAA20262.1; --  
 DR GO; GO:0003723; F:RNA binding; IEA.  
 DR GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.  
 DR GO; GO:0006350; P:transcription; IEA.  
 DR InterPro; IPR002581; Morbilli\_P.  
 DR Pfam; PF01647; Morbilli\_P; 1.  
 FT NON\_TER 128 128  
 SQ SEQUENCE 128 AA; 13308 MW; CF2A0AFC3673F78 CRC64;

Query Match 100.0%; Score 26; DB 12; Length 128;  
 Best Local Similarity 100.0%; Pred. No. 3.5e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EKAGG 5  
 |||||  
 Db 53 EKAGG 57

Search completed: September 18, 2004, 04:29:49  
Job time : 24.7451 secs

---